Sensemaking in Immersive Space to Think: Exploring Evolution, Expertise, Familiarity, and Organizational Strategies

Kylie Marie Davidson

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 and Applications

Chris L. North, Chair
Doug A. Bowman, Co-chair
Kirsten Whitley
Nicholas F. Polys
Brendan M. David-John

August 6th, 2024
Blacksburg, Virginia

Keywords: Human-Computer Interaction, Immersive Analytics, Sensemaking, Visualization, Distributed Cognition
Copyright 2024, Kylie Marie Davidson

Sensemaking in Immersive Space to Think: Exploring Evolution, Expertise, Familiarity, and Organizational Strategies

Kylie Marie Davidson

(ABSTRACT)

Sensemaking is the way in which we understand the world around us. Pirolli and Card developed a sensemaking model related to intelligence analysis, which involves taking raw, unstructured data, analyzing it, and presenting a report of the findings. With lower-cost immersive technologies becoming more popular, new opportunities exist to leverage embodied and distributed cognition to better support sensemaking by providing vast, immersive space for creating meaningful schemas (organizational structures) during an analysis task. This work builds on prior work in immersive analytics on the concept of Immersive Space to Think (IST), which provides analysts with immersive space to physically navigate and use to organize information during a sensemaking task. In this work, we performed several studies that aimed to understand how IST supports sensemaking and how we can develop additional features to better aid analysts while they complete sensemaking in immersive analytics systems, focusing on non-quantitative data analysis. In a series of exploratory user studies, we aimed to understand how users' sensemaking process evolves during multiple session analyses, which identified how the participants refined their use of the immersive space into later stages of the sensemaking process. Another exploratory user study highlighted how professional analysts and novice users share many similarities in immersive analytic tool usage during sensemaking within IST. In addition to looking at multi-session analysis tasks, we also explored how sensemaking strategies change as users become more familiar with the immersive analytics tool usage in an exploratory study that utilized multiple analysis tasks completed over a series of three user study sessions. Lastly, we conducted a comparative user study to evaluate how the addition of new organizational features, clustering, and linking affect sensemaking within IST. Overall, our studies expanded the IST tool set and gathered an enhanced understanding of how immersive space is utilized during analysis tasks within IST.

Sensemaking in Immersive Space to Think: Exploring Evolution, Expertise, Familiarity, and Organizational Strategies

Kylie Marie Davidson

(GENERAL AUDIENCE ABSTRACT)

Sensemaking is a process we do in our daily lives. It is how we understand the world around us, make decisions, and complete complex analyses, like journalists writing stories or detectives solving cases. Sensemaking involves gathering information, making sense of it, developing hypotheses, and drawing conclusions, similar to writing a report. This work builds on prior work in Immersive Space to Think (IST), which is a concept of using immersive technologies (Virtual /Augmented Reality) to support sensemaking by providing vast 3D space for organizing the data used in a sensemaking task. Additionally, using these technologies to support sensemaking provides benefits such as increased space for analysis, increased engagement, and natural user interaction, which allow us to interact with information used during sensemaking tasks in new ways. In IST, users are able to move virtual documents around in the space around them to support their analysis process. In this work, we ran a study focused on multi-session analysis within IST, revealing how users refined their document placements over time while completing sensemaking tasks within IST. We also ran a study to understand how professional analysts' and novice users' analysis with IST differed in the IST tool usage. In another user study, we explored how users' strategies for sensemaking and document layouts changed as they became more familiar with the IST tool. Lastly, we conducted a comparative user study to evaluate how new features like clustering and linking affected analysis within IST. Overall, our work contributed to an enhanced understanding of how immersive space is utilized during analysis tasks within IST.

Acknowledgments

Personal Acknowledgements

To my mentors To my advisors, Dr. Chris North and Dr. Doug Bowman, I am deeply grateful for your unwavering support and guidance. Dr. Chris North, you welcomed me into your lab early on, helping me get started in research with constant encouragement. Your high-level insights and unique perspective on my results have been invaluable, allowing me to think critically and deeply about my work. Beyond the research, your sense of humor and ability to keep meetings light-hearted created a positive and enjoyable working environment. The spring and fall cookouts at your home were wonderful opportunities to bond with the lab and build a sense of community. Dr. Doug Bowman, your leadership has shown me how to be a better researcher. Your thoughtful comments and meticulous attention to detail have ensured that I consider my work from multiple perspectives, guiding me to produce my best work. The meaningful structure you provided helped me stay on track and focused as a student. You consistently encouraged me to think about the bigger picture and how my research fits into a broader context. Your dedication and willingness to provide deep feedback or provide words of wisdom have been truly inspiring. The balance of the complementary strengths you both provided has been instrumental in my development. I am profoundly thankful for your mentorship. Dr. Kirsten Whitley, in many ways, you acted as a third advisor to me, becoming deeply involved in my research and supporting me along the way. You facilitated amazing opportunities, such as presenting my work to key figures and engaging in impactful projects, which helped me to grow as a researcher. Thank you for your constant encouragement and support throughout this journey. Dr. Brendan David-John and Dr. Nicholas Polys, your encouragement, insights, and thoughtful comments have greatly contributed to my development. I am thankful for your time and efforts in teaching and mentoring me. To my undergraduate professors, Dr. Dee Weikle, Dr. John Bowers, Dr. Chris Mayfield, Dr. Kevin Molloy, and Dr. Michael Stewart, thank you for inspiring me to pursue graduate school and for your continued guidance during my time at JMU and VT. Thank you all for your open-door policy, your willingness to answer my questions and your patience. Your mentorship laid the foundation for my academic journey.

To the Center for Human-Computer Interaction and New Horizons From my first week at Virginia Tech, CHCI has impacted my journey here. Thank you to this community for providing me with a place that made this large campus feel small and for the amazing research development opportunities. Thank you to New Horizons for the friends, mentors, and professional development opportunities, and a special thank you to Renee Cloyd for the vibrant community you built within New Horizons, your mentorship, and the weekly Lunch & Learn, which always provided friendly faces.

To my friends and colleagues The InfoVis and 3DI labs, there are so many of you who have impacted my journey. Thank you to my lab mates for always being willing to answer questions, share laughs, and guide me through this process. To the IST Team, Dr. Lee Lisle, and Ibrahim Tahmid, thank you for your help, collaboration, encouragement, and countless moments of camaraderie. To my friends, thank you all for your constant support and for checking in on me.

To my family A short acknowledgments section is not enough space to properly thank you all for your never-ending support and belief in me. Mom and Dad, thank you for everything—from renovating the loft to all our phone calls while I walked to and from school, your encouragement to pursue computer science as an undergrad, and your helping me believe in myself when needed. I can never properly thank you for all the love and support you've provided throughout this journey. Darcy, thank you for making grad school less lonely with our family dinners, dance parties, and countless adventures, such as rock climbing, camping, golfing, kayaking, hiking, or even binging a Netflix show too late into the morning. Your love and support as my big sister have been a constant comfort and inspiration throughout my journey. Seeing you pave the way with your achievements and earning your PhD in biochemistry gave me the belief that I could pursue mine. Thank you for being a guiding light in my academic journey. Trevor, thank you for all the phone calls to chit-chat while you drove home from golf or work. Staying connected with you, even from four hours away, meant so much to me. Betsy, thank you for supporting my journey to a PhD even before I started grad school, being there for all the milestones, and celebrating them along the way. Aidan, thank you for always helping with math concepts I was unfamiliar with, assisting with editing papers, and being part of the "Davidson South" family. Emily, thank you for your positive support and encouragement and for always including us in party invites, even when you suspected we might be unable to make it. Eddie, your unwavering support since before I applied to grad school has been incredible. Thank you for supporting my dream of getting a PhD, for being there through all the ups and downs, for the much-needed pep talks, and for celebrating all the milestones along the way. Your love and belief in me have been a constant source of strength and motivation. I love you.

Funding Sources

This research was partially supported by NSF I/UCRC CNS-1822080 via the NSF Center for Space, High-performance, and Resilient Computing (SHREC), the Laboratory for Analytic Sciences (LAS), and the Dr. Richard E. Nance Graduate Fellowship - Virginia Tech Department of Computer Science.

Attribution

First, we want to acknowledge that the Immersive Space To Think concept has been a

large project involving multiple students, faculty members, and collaborators. Individuals who have contributed to the implementation of the various IST prototypes include Dr. Lee Lisle, Ibrahim A. Tahmid, Payel Bandyopadhyay, Kyle Tanous, Dr. Edward J.K. Gitre, Dr. Kirsten Whitley, Dr. Chris North, and Dr. Doug A. Bowman.

Chapter 3 is based on the paper titled "Exploring the Evolution of Sensemaking Strategies in Immersive Space to Think," which was accepted for publication at IEEE TVCG 2023 ©2023 IEEE. Reprinted, with permission, from [19]. This work was conducted in collaboration with Dr. Lee Lisle, Dr. Kirsten Whitley, Dr. Doug Bowman, and Dr. Chris North.

Chapter 4 was conducted in collaboration with Dr. Lee Lisle, Ibrahim A. Tahmid, Dr. Kirsten Whitley, Dr. Chris North, and Dr. Doug Bowman.

Chapter 5 is based on the paper titled "Uncovering Best Practices in Immersive Space to Think," which was presented at IEEE ISMAR 2023. Reprinted, with permission, from [20]. This work was conducted in collaboration with Dr. Lee Lisle, Ibrahim A. Tahmid, Dr. Kirsten Whitley, Dr. Chris North, and Dr. Doug Bowman.

Chapter 6 was conducted in collaboration with Dr. Lee Lisle, Ibrahim A. Tahmid, Dr. Kirsten Whitley, Dr. Chris North, and Dr. Doug Bowman.

Contents

List of Figures xiii

List of Tables xx

1 Introduction 1

1.1 Background & Motivation . . . 1

1.2 Immersive Space to Think . . . 3

1.2.1 Features . . . 4

1.2.2 Prior work in Immersive Space to Think . . . 7

1.3 Research Questions . . . 10

1.4 Structure of Dissertation . . . 11

2 Review of Literature 12

2.1 Foundations in Human-Computer Interaction . . . 12

2.1.1 Embodied and Distributed Cognition . . . 12

2.1.2 Sensemaking . . . 13

2.2 Visual Analytics . . . 15

2.2.1 Definition & Model . . . 15

2.2.2 Visual Analytics Examples . . . 16

2.3 Immersive Analytics . . . 18

2.3.1 Immersive Technologies . . . 18

2.3.2 Immersive Analytics Defined . . . 19

2.3.3 Large Workspaces in Immersive Analytics . . . 21

3 Exploring the Evolution of Sensemaking in Immersive Space to Think 23

3.1 Introduction . . . 23

3.2 Immersive Space to Think Prototype . . . 25
3.2.1 New Interaction Techniques . . . 25
3.3 User Study . . . 25
3.3.1 Goals and Research Questions . . . 25
3.3.2 Apparatus . . . 26
3.3.3 Experimental Task . . . 26
3.3.4 The Dataset . . . 28
3.3.5 Participants . . . 28
3.3.6 Measures . . . 29
3.3.7 Procedure . . . 29
3.4 Results and Discussion . . . 31
3.4.1 RQ1.1A: Spatial Structures and Change Over Time . . . 32
3.4.2 RQ1.1B: Mapping of Spatial Layout to Report Structure . . . 39
3.4.3 RQ1.2: Physical Navigation . . . 40
3.4.4 RQ1.3: Bridging Technologies . . . 43
3.4.5 RQ1.4: Report Scores and Strategies . . . 43
3.5 Limitations . . . 47
3.6 Conclusions . . . 48

4 Investigating Professional Intelligence Analyst Sensemaking in Immersive Space to Think 49

4.1 Introduction . . . 49
4.2 User Study . . . 51
4.2.1 Goals and Research Questions . . . 51
4.2.2 Apparatus . . . 52
4.2.3 Experimental Task . . . 53
4.2.4 Dataset . . . 54
4.2.5 Participants . . . 54
4.2.6 Measures . . . 55

4.2.7 Procedure . . . 55
4.3 Results and Discussion . . . 57
4.3.1 RQ 1.5: Spatial Layouts and Change Over Time . . . 57
4.3.2 RQ 1.6: Mapping of Spatial layouts to Report Structure . . . 64
4.3.3 RQ 1.7: Physical Navigation . . . 65
4.3.4 RQ 1.8: Strategies and Report Scores . . . 69
4.3.5 Key Takeaways . . . 72
4.4 Limitations . . . 73
4.5 Conclusions . . . 74

5 Uncovering Best Practices in Immersive Space to Think 76
5.1 Introduction . . . 76
5.1.1 Immersive Analytics System Usage and Learning . . . 77
5.2 Immersive Space to Think Approach . . . 78
5.3 User Study . . . 78
5.3.1 Goals and Research Questions . . . 78
5.3.2 Apparatus . . . 79
5.3.3 Datasets . . . 80
5.3.4 Experimental Tasks . . . 80
5.3.5 Participants . . . 81
5.3.6 Measures . . . 81
5.3.7 Procedure . . . 82
5.4 Results and Discussion . . . 82
5.4.1 RQ2.1: Spatial Layouts and Deeper Organization . . . 83
5.4.2 RQ2.2: Strategies for Analysis . . . 88
5.4.3 RQ2.3: Correlations with Sensemaking Quality . . . 90
5.4.4 IST Analysis Best Practices . . . 92
5.5 Limitations . . . 93
5.6 Conclusions . . . 93

6 Designing for Best Practices in Immersive Space to Think **95**

6.1 Feature Design . . . 96

6.1.1 Initial Prototyping . . . 96

6.1.2 Expert Feedback Sessions . . . 105

6.1.3 Results from Feedback Sessions . . . 106

6.2 User Study Conditions and Experimental Task . . . 108

6.2.1 Task Comparison . . . 110

6.3 New Organizational Features Iterative Design . . . 117

6.3.1 IST Document Modifications . . . 117

6.3.2 Clustering Modifications . . . 118

6.3.3 Manual Linking Modifications . . . 119

6.3.4 Discovery Linking Modifications . . . 119

6.3.5 Other Modifications for Experimental Task . . . 121

6.4 User Study . . . 122

6.4.1 Goals and Research Question . . . 122

6.4.2 Apparatus . . . 123

6.4.3 Experimental Task and Conditions . . . 124

6.4.4 Participants . . . 124

6.4.5 Measures . . . 125

6.4.6 Procedure . . . 125

6.4.7 Results & Discussion . . . 127

6.4.8 Key Takeaways . . . 149

6.5 Limitations & Future Work . . . 150

6.6 Conclusions . . . 151

7 Conclusions & Future Work **152**

7.1 Conclusions . . . 152

7.1.1 RQ 1: How do sensemaking strategies evolve during multi-session analysis in Immersive Space to Think? . . . 152

7.1.2 RQ 2: How does familiarity affect initial strategies and layouts used for immersive sensemaking? . . . 153
7.1.3 RQ 3: How can we support best practices for immersive sensemaking with additional organizational features? . . . 153
7.2 Future Work Directions . . . 154
7.2.1 Scalability . . . 154
7.2.2 Longitudinal Studies . . . 154
7.2.3 Collaboration . . . 155
7.2.4 Human-Machine Teaming . . . 156

Bibliography **157**

Appendices **167**

Appendix A Exploring the Evolution of Sensemaking Strategies in Immersive Space to Think **168**
A.1 Informed Consent Form . . . 168
A.2 Questionnaires & Interviews . . . 173
A.2.1 Pre-Study Questionnaire . . . 173
A.2.2 Session 1 Interview . . . 176
A.2.3 Session 2 Interview . . . 176
A.2.4 Session 3 Interview . . . 176
A.3 Post-Study Semi-Structured Interviews . . . 177
A.4 COVID-19 Consent Addendum . . . 177
A.5 IRB Approval Letter . . . 179

Appendix B Exploring Professional Analyst Sensemaking Strategies in Immersive Space to Think **182**
B.1 Informed Consent Form . . . 182
B.2 Questionnaires & Interviews . . . 186
B.2.1 Pre-Study Questionnaire . . . 186

B.2.2 Condition 1 Interviews . . . 189
B.2.3 Condition 2 Interviews . . . 190
B.3 Post-Study Semi-Structured Interviews . . . 191
B.3.1 Condition 1 . . . 191
B.3.2 Condition 2 . . . 191
B.4 IRB Approval Letter . . . 192

Appendix C Gathering Expert Feedback on Deeper Organizational Features 195
C.1 Informed Consent Form . . . 195
C.2 Questionnaires & Interviews . . . 199
C.2.1 Pre-Study Questionnaire . . . 199
C.3 Post-Study Semi-Structured Interviews . . . 202
C.4 IRB Approval Letter . . . 203

Appendix D Evaluating Organizational Feautres in Immersive Space to Think 206
D.1 Informed Consent Form . . . 206
D.2 Questionnaires & Interviews . . . 211
D.2.1 Pre-Study Questionnaire . . . 211
D.3 Post-Task Semi-Structured Interviews . . . 212
D.4 Post-Study Semi-Structured Interviews . . . 212
D.5 IRB Approval Letter . . . 213

Appendix E Report Grading Rubrics 216

List of Figures

1.1 The Sensemaking Process as Defined by Pirolli & Card [73]. The process contains a series of steps used while an analyst develops a hypothesis or finding from unstructured data. This bidirectional process consists of two sub-loops - the foraging loop and the sensemaking loop. The goal at the end of the sensemaking process is to present findings. 2

1.2 Left: Bulletin Board loaded with documents for analysis. Right: Example space with documents organized within the 3D space, Annotations (Green), Labels (Orange), and Notes (Purple). 4

1.3 Left: Keyboard stand with wireless keyboard, HTC Vive Tracker, and example space for a controller. The keyboard is on a wheeled table, allowing the participant to move it around the space as needed during their analysis. Right: Virtual representation of the keyboard. Attached to the keyboard are four menu options. These options from left to right are New Note, New Label, Search, and Save. These menu options are attached to the table and will move with the table as it is moved around the space. For text entry, the participant could view their fingers on the keyboard while peaking out of the bottom of the headset enabled by small amounts of foam on the headset being cut out. 5

1.4 Example document provided for analysis with the User-Interface Panel, which is activated on hover on the Left Hand Side. From top to bottom, the buttons are Highlight, Copy to Clipboard, Copy to Clipboard with Citation, and Create Attached Note (Annotation). 5

2.1 Klein & Moon Data Frame Sensemaking Model. This model focuses on creating mental model frames of the data where these frames are constantly refined, adjusted, and elaborated. 14

2.2 Sacha et al.'s Knowledge Generation Model for Visual Analytics. The model consists of two parts: the computer and the human. In this model, the computer or left-hand side illustrates the visual analytic system and the right-hand side indicates the knowledge generation process. The right-hand side has three sub-loops - exploration, verification, and knowledge generation. This model highlights the human-machine teaming needed in visual analytic systems. 15

2.3 The **Space to Think** or **Analyst Workstation** presented by Andrews et al. It was made up of an array of 4 x 2 30" monitors that were slightly curved around the user [1]. 17

3.1 Overview of an Immersive Space to Think Analysis Space. This example is of participant five's spatial layout over time from left to right: start, end of session 1, end of session 2, and end of session 3. Top Row: view of immersive space. Bottom Row: view of 3D scatterplots used for analysis in a top-down orientation. 24

3.2 Left: Keyboard stand with wireless keyboard, HTC Vive Tracker, and example space for a controller. The keyboard is on a wheeled table which allows the participant to move the keyboard around the space as needed during their analysis. Right: Virtual representation of the keyboard. Attached to the keyboard are four menu options. These options from left to right are *New Note, New Label, Search, and Save*. These menu options are attached to the table and will move with the table as it is moved around the space. 27

3.3 All participants' final spatial layouts formed during their analysis. These spatial layouts are portrayed using a top-down scatter plot. The red area on the left side denotes the bulletin board. Each document is represented using a blue dot, while notes are red and labels are yellow. The red curves approximate the overall organization structures formed by the participants. 33

3.4 Number of documents moved during each session of analysis (Left). Distances (meters) that participants moved each document in each session (Right). . . 34

3.5 Evolving spatial organization for participant 4. From left to right we have the start, middle, and end (session 1) middle, end (session 2) and middle, end (session 3). 34

3.6 Reporting agency of each of the documents in P4's spatial layout after session 1. Yellow represents CIA, Blue represents FBI, Green represents SI, and gray is the prompt/note. 37

3.7 Two columns that are organized into timeline order, on the left the top document is posted on April 18, and bottom on April 22. The right-hand column goes from April 24 (top) to April 27 (bottom) 38

3.8 3D scatterplot of P6's spatial layout with the document reference order traced. Gray documents were not referenced in the report. The red rectangle represents the bulletin board, and the open gray circle represents a free-standing note. Colored documents were referenced in the report coordinated with the paragraph coloring on the right. 40

3.9 Participant walking movement normalized by time spent in headset per session. 41

3.10 Trajectory of movement for P1 during each session (top row) and the dwell times of the participant across each session (bottom row) where a darker color indicates more time spent in that location. Each plot has the document locations at the end of the session represented with blue (document), yellow (label), and red (note). The red line denotes the bulletin board. 42

3.11 Report Correctness score vs. Report Quality score. 44

3.12 P5's Spatial Layouts: end of session one (left), and end of session two (right). The colors represent plotline encodings where red represents distractors, blue/-green/purple represents plotlines, yellow represents coordination, and black represents notes/labels/prompt. 45

3.13 P8's spatial layout after the end of session two. Color encoding is the same as figure 3.12. 46

4.1 High-Level View of an Immersive Space to Think Analysis. This example is of Participant Five's analysis over time. From left to right: start, middle of session 1, end of session 1, middle of session 2, end of session 2, middle of session 3, end of session 3. Top Row: Top-down orientation into the immersive space. Bottom Row: top-down view of 3D scatter plots used for data analysis in this work. The blue dots represent the documents provided to the users for analysis, and the red dots represent notes/labels created by the user during analysis. 50

4.2 Left: Text Entry System. A HTC Vive Tracker on the wheeled table, a wireless keyboard, and space for the controller to be set down while typing. The keyboard was replaced with a 2019 Mac Book Pro using Google Docs during report writing. Right: VR Headset view of the tracked keyboard and table. There are four menu buttons attached to the keyboard. These buttons are New Note, New Label, Search, and Save. 52

4.3 Top-Down View at High-Level Organizational Layout of the analysis spaces. Left: Semi-Cylindrical Layout, Middle: Cylindrical Layout, Right: Environmental Layout. The white documents are those provided for analysis, and the blue documents indicate a note or label created during the analysis. In the bottom right-hand corner is a copy of the task as detailed in section 4.2.3. 58

4.4 Deeper Organizational Features found in the analysis spaces. Left: Column/-Timeline organization with additional labeling of people of interest within the columns, Middle: Scratch Space for the analyst's comments, Right: Network Graph created with the label feature. 59

4.5 From left to right, End of Session One, End of Session Two, and End of Session Three of P4's analysis space. The top row is a top-down view into the space, and the bottom row is a rotated view looking towards the bulletin board. The colors in this image indicate the clustering applied to each document by the OPTICS algorithm. The black documents were "noise" detected by the OPTICS algorithm. 62

4.6 The total number of documents moved was significantly different from session 1 to 3 ($p = 0.004$) and again from session 2 to 3 ($p = 0.007$). There are also significant differences in the average distance a document moved from sessions 1 to 3 ($p = 0.0007$) using an Each Pair Students T-Test with Bonferroni's correction. 63

4.7 Left: P3 spatial layout with a reference line drawn through it. The green mesh indicates the bulletin board. The red, opaque documents are cited in paragraph one, the yellow documents in paragraph 2, the blue in paragraph 3, and the pink in paragraph 4 of the report. Right: P5 reference order of moving relatively top-down through the columns within their analysis space for citing documents. The zoomed portion is oriented as if the participant is standing in the center of the analysis space, staring at documents. Most Red, Orange, and Yellow documents are on the top of the columns, with Pinks, and Purples and Blues toward the bottom of the columns. In this figure, the documents are cited in paragraph order Red, Orange, Yellow, Green, Blue, Indigo, Violet, Pink (ROYGBIV + Pink). 65

4.8 Movement data normalized by time in the headset. We can see a decrease in movement over time with significant differences between sessions 1 and 3 ($p = 0.02$) and 2 and 3 ($p = 0.03$) using an Each Pair Students T-Test with Bonferroni's correction. 66

4.9 From left to right, session 1, session 2, and session 3 for P6's (top) and P11's (bottom) analysis. The red plane represents the bulletin board, the blue points indicate documents, the red colored points are notes created by the user, and the yellow points are labels created by the user. 67

4.10 From left to right, session 1, session 2, and session 3 for P6's (top) and P11's (bottom) analysis. The red plane represents the bulletin board, the blue points indicate documents, the red colored points are notes created by the user, and the yellow points are labels created by the user. The heat maps show the time spent (dwell) within a one sq ft area of the tracked space. . . 68

4.11 Left - Highest scorers, Right - Lowest Scorers. The x-axis represents time within each of the sessions from the start of the analysis to the end of the analysis. The y-axis is the frequency of each interaction normalized between 0 and 1. 71

5.1 A top-down view into the final spatial layout of a participant. From left to right, session 1, session 2, and session 3. We can see distinct differences in the use of space and deeper organizational features used. The document located in the bottom right corner is the analysis prompt. 77

5.2 From left to right, Semi-Cylindrical, Environmental, and Cylindrical layouts created by P13, P7, and P3, respectively. The red mesh represents the virtual bulletin board, red dots represent notes, yellow dots represent labels, and blue dots represent the documents used during analysis. 83

5.3 From left to right, session 1, session 2, and session 3 angular placement of documents. The direction towards the bulletin board is at the top of the circle. The space is broken into 15-degree bins, and each bar represents the total count of documents across all participants in that 15-degree slice of the space. 85

5.4 From left to right, session 1, session 2, and session 3 clustering using OPTICS for P11's final spatial layout. The red plane represents the bulletin board, the black points indicate documents not assigned to a cluster ("noise"), and the colored points indicate the clusters to which the documents were assigned. 86

5.5 Total count of strategy usage across the sessions. Blue is session 1, red is session 2, and green is session 3. 89

6.1 Cluster Creation Overlap Technique 97

6.2 An IST view of the cluster alignment feature with the Row alignment applied. The menu for selecting a quick alignment can be seen in the upper left-hand corner of the cluster. The options from top to bottom are Row, Column, and Timeline. 99

6.3 An example cluster starting point for showing the different alignment options. 99

6.4 Row Alignment Cluster Example 100

6.5 Column Alignment Cluster Example 101

6.6 Example of a Manual Link Created During Analysis. 103

6.7 Example of a Discovery Link Executed on the name "Faysal Goba". 104

6.8 Connections Game Example 110

6.9 The Sensemaking Process as Defined by Pirolli & Card [73]. The process contains a series of steps used while an analyst develops a hypothesis or finding from unstructured data. Rectangular boxes represent the approximate data flow and the circles represent process flow during the sensemaking process. . 111

6.10 An example Card for the Connections Based Game. The document had two arrow buttons to access the word's definitions and synonyms. When the next page button is pressed, the card is scrolled through the definition and synonym lists. When the definitions or synonyms are displayed, a title bar appears over the document with the document title. 117

6.11 Modifications made to the clustering features. Left: Anchor points were added to the cluster container (top, right, bottom, left), which can be used to link information to clusters or link clusters to clusters. Middle: User Interface panel in the top left corner for changing the clusters' color with preset options of green, yellow, red, and blue. Right: Added the ability to place labels directly into clusters. 118

6.12 The process of deleting a link. First, the user hovers over the unwanted link using the ray, and then the user clicks the trigger on the controller, resulting in the link disappearing, right. 119

6.13 Radial Menu for changing the color of the links. The menu is opened when the user presses down on the controller trackpad. Then, the user can move their thumb around the track pad, which moves the cursor (black dot). Whenever the user releases their thumb, that color becomes the new link color. For example, if this user released their thumb, the new link color would be set to green. 119

6.14 Discovery Linking Feature Example. Left - the button used to open a discovery link, Middle-Left - An open discovery link (purple links) on the work AUDITORY. Middle-Right - State after the check box (keep) button is pressed on SONIC. Right - After the discovery link is toggled off, SONIC is kept around, and ACOUSTIC is returned to its previous location. 120

6.15 Keyboard Table used for Connections-Based task. The table had three buttons attached to it: Submit, New Label, and Search. 121

6.16 Breakdown of the (Total Distances Between Document / Number of Document). The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05, ** $\leq$ 0.001 128

6.17 Breakdown of the Silhouette Scores. The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05, ** $\leq$ 0.001 . . . 129

6.18 Breakdown of the CH Scores. The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05, ** $\leq$ 0.001 130

6.19 Average Total Number of Grabs Across Participants (left), Definition/Synonym Access (middle), and Search (right) interactions by condition. The vertical black line annotation shows the 95% confidence interval. Significance is denoted with * $\leq$ 0.05, ** $\leq$ 0.001 133

6.20 NASA TLX Result Box plots broken down by Condition. The dotted line and diamond visualize the mean and Standard Deviation. Higher scores represent a higher workload. Whiskers show the spread of the data without outliers. . 137

6.21 NASA TLX Result Box plots broken down by Presence of Clustering or Linking. Higher scores represent a higher workload. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05 138

6.22 Results from special questions on the post-task questionnaire. Top - Condition specific breakdown, bottom - Clustering and Linking Presence breakdown. The scale is Strongly Disagree (Lower Scores) and Strongly Agree (Higher Scores). Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05 and ** $\leq$ 0.001 139

6.23 UEQ Results Box plots broken down by Condition. The dotted line and diamond visualize the mean and Standard Deviation. Higher scores indicate better usability. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05 140

6.24 UEQ Results Box plots broken down by Present of Clustering or Linking. Higher scores indicate better usability. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05 141

6.25 Results from special questions on confidence in solution on post-task questionnaire. The scale is Strongly Disagree (Lower Scores) and Strongly Agree (Higher Scores). Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05. 145

6.26 Task Performance Breakdown for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers. 100 represents a perfect score on the task. 146

6.27 Count of correct groups of four breakdowns for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers. 147

6.28 The time taken to submit the solution for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05 147

List of Tables

5.1 Final layout categorizations for the end of each sensemaking session 84

5.2 Average score and standard deviation for each strategy 91

6.1 Best Practices and How to Design to Support Them 96

6.2 Table summarizing the comparison between Sign of the Crescent and the Modified Connections Game when evaluated on the Approximate Data Flow stages of the Sensemaking Process. 114

6.3 Table summarizing the comparison between Sign of the Crescent and the Modifieid Connections Game when evaluated on the Approximate Process Flow stages of the Sensemaking Process. 116

6.4 Modified Connections Task Scoring Rubric. 145

Chapter 1

Introduction

1.1 Background & Motivation

Sensemaking is a process woven into our daily routines, guiding decisions ranging from selecting what to wear to deciding what to cook for dinner. At its core, sensemaking can be defined as "how people make sense out of their experiences in the world" [21]. This cognitive process involves synthesizing various streams of information, allowing us to interpret and make informed decisions. The spectrum of sensemaking tasks spans from straightforward choices like what to wear or cook, taking mere minutes, to intricate endeavors such as crime investigations, which unfold over days, months, or even years. Professions that center around sensemaking, such as journalism [15], historical interpretation [51], crime investigation [41], and intelligence analysis [73], exemplify the diverse domains where this cognitive process plays a pivotal role.

Pirolli & Card proposed a model for the sensemaking task, as depicted in Figure 1.1, consisting of a series of iterative steps used by analysts to develop a hypothesis or finding from unstructured raw data. The iterative steps within this model form two sub-loops: **Foraging Loop** and **Sensemaking Loop**. The Foraging sub-loop is primarily dedicated to evidence gathering, involving activities such as data "Search & Filter," "Read & Extract" for meaning, and "Search for Relations." On the other hand, the Sensemaking sub-loop focuses on formulating hypotheses and preparing findings. This phase encompasses actions like "Search for Evidence and Support," "Develop Schemas," "Reevaluate" findings, and the preparation of a "Presentation." In intelligence analysis, the presentation can be a written briefing of their findings. Analysts engage with diverse unstructured data throughout the sensemaking process, constructing a mental model or schema to interpret its meaning, formulating hypotheses, and ultimately compiling a detailed report of their findings. Importantly, this process is bidirectional, enabling analysts to navigate both upward and downward within the sensemaking model as they progress toward the ultimate goal of developing a comprehensive presentation of their findings.

This cognitive process imposes demands on analysts, with the level of effort and structure increasing as they approach the presentation stage. Moreover, the consequences of sensemaking become particularly high-stakes in fields like crime investigations and intelligence analysis, where decisions hold critical implications for trials or national security. Delving into the vast realm of big data [79] introduces additional layers of complexity, intensifying

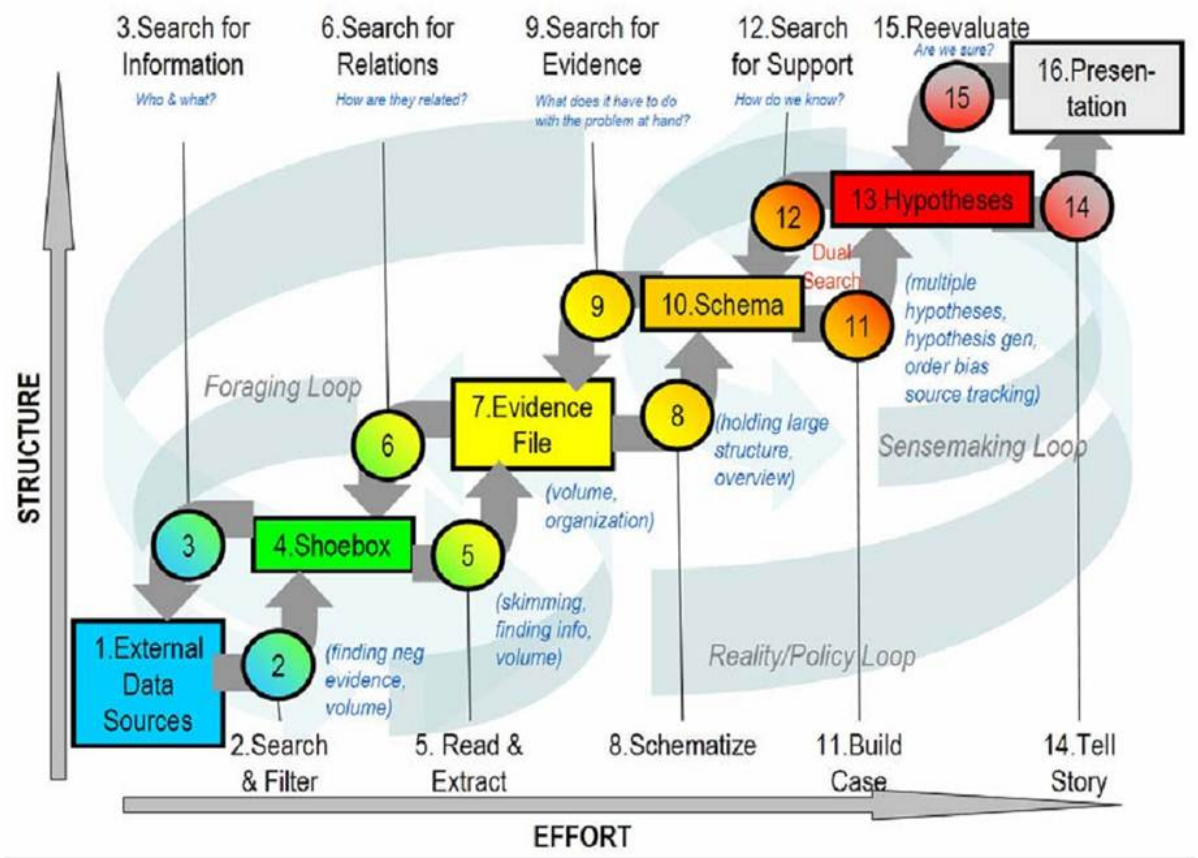


Figure 1.1: The Sensemaking Process as Defined by Pirolli & Card [73]. The process contains a series of steps used while an analyst develops a hypothesis or finding from unstructured data. This bidirectional process consists of two sub-loops - the foraging loop and the sensemaking loop. The goal at the end of the sensemaking process is to present findings.

the challenges confronted by journalists, crime investigators, and intelligence analysts. Their roles become increasingly overwhelming in this evolving landscape due to the sheer volume of data requiring gathering and processing.

In efforts to enhance sensemaking around big data, many technologies have been developed to help support analysts during sensemaking. One notable research area addressing this challenge is visual analytics, defined as the "process of analytical reasoning facilitated by interactive visual interfaces" [89]. Its goal is to synthesize information, detect expected and unexpected patterns, and offer timely support [49]. One approach within visual analytics involves providing interactive visualization assistance during the sensemaking process [7, 26, 86]. This approach allows for extracting and linking essential information through interactive diagrams. Another approach in visual analytics is to provide large high-resolution space for analysis [1, 12, 33]. This approach leverages the surrounding space as a form of externalized memory, enhancing the sensemaking process for analysts by providing more space to complete an analysis task.

However, in recent years, the emergence of lower-cost Virtual Reality/ Augmented Reality (VR/AR) immersive display technologies has created new opportunities to explore how immersive space can be used to assist in sensemaking tasks. Immersive Space to Think is foundational in this exploration, which builds upon the concept of offering expansive physical spaces for analysis or sensemaking tasks, as established in prior work [1].

1.2 Immersive Space to Think

Immersive Space to Think (IST) is a concept of using VR/AR technologies to provide users with a large immersive tracked area to use while completing a sensemaking task [5, 56, 59]. IST goes beyond the original space to think, enabling users to externalize their cognition into the 3D environment by offloading cognition from their analysis into the immersive space around them through the organizations and structures created during their analysis. Additionally, IST leverages the benefits of immersive technologies by providing richer sensory information, including better depth cues, increased spatial orientation understanding, expansive 360 degrees of space, 6 degrees of freedom for interaction, better peripheral awareness, and increased engagement compared to traditional desktop displays [1, 10, 68]. In IST, users are able to pull documents off of a virtual bulletin board, as seen in figure 1.2, left, and organize them in the environment around them, as seen in figure 1.2, right. Additionally, as the users are completing analysis within the system, they physically navigate (e.g., walk) around the analysis space.

As IST is a concept, many iterations of its design, features, and hardware have been used to achieve immersion. Current versions of IST are developed using an immersive Virtual or Augmented Reality headset. The users then interact with the system using a handheld controller in their dominant hand. Then, to provide tracking, the SteamVR lighthouse

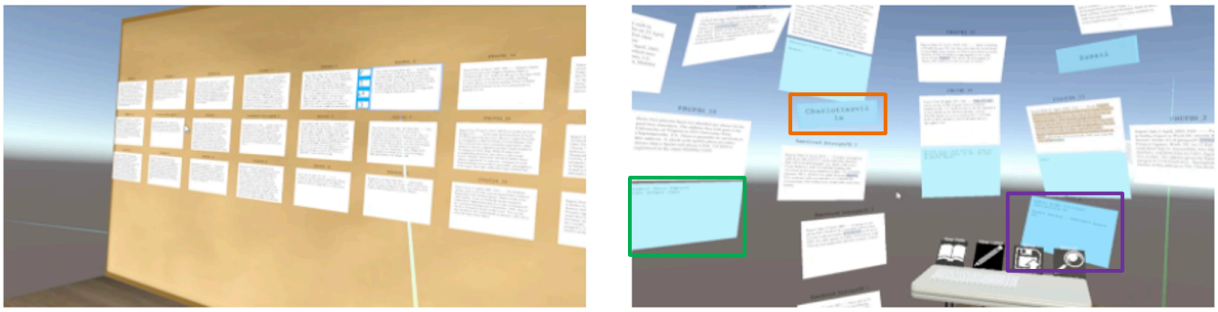


Figure 1.2: Left: Bulletin Board loaded with documents for analysis. Right: Example space with documents organized within the 3D space, Annotations (Green), Labels (Orange), and Notes (Purple).

tracking system is used.

1.2.1 Features

With this research focusing on text-based sensemaking, we wanted to provide features for completing this style of analysis. Below, we detail the basic set of features that are supported in IST.

Text Entry

In order to support the offloading of analyst comments into the environment around the user and the entire sensemaking process up through presentation, we needed to develop support for text entry within IST. Using a tracked physical rolling table and keyboard, as seen in figure 1.3, participants can enter text with a Bluetooth-connected keyboard and can move the keyboard around the immersive environment with them. The table and keyboard are represented with similarly sized 3D models in the virtual environment. In chapters 3, 4, and 5, as the user was in VR and the user's hands were not tracked, the user must place their hands properly on the keyboard using the sense of touch or by viewing their physical hands and keyboard outside of the headset. In chapter 6, the users were using AR video-pass through so the participants could see their finger on the keyboard in close to real-time. In IST, the system has three primary forms of text entry: annotations (green box), notes (purple box), and labels (orange box). All forms of these can be seen in figure 1.2, right, represented with a blue background color.

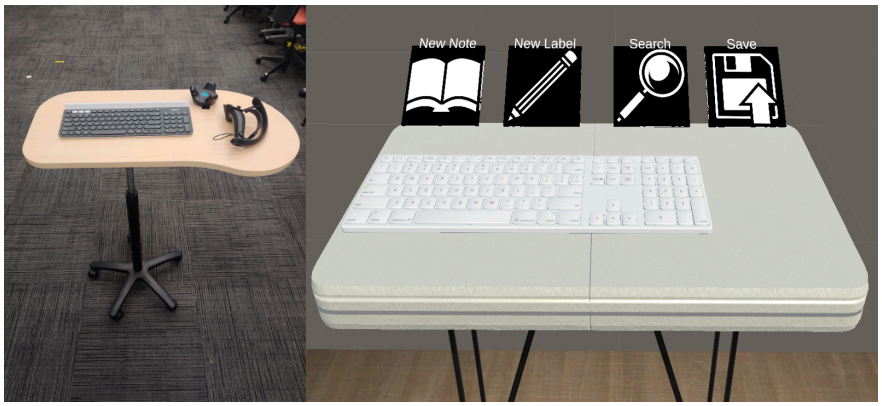


Figure 1.3: Left: Keyboard stand with wireless keyboard, HTC Vive Tracker, and example space for a controller. The keyboard is on a wheeled table, allowing the participant to move it around the space as needed during their analysis. Right: Virtual representation of the keyboard. Attached to the keyboard are four menu options. These options from left to right are New Note, New Label, Search, and Save. These menu options are attached to the table and will move with the table as it is moved around the space. For text entry, the participant could view their fingers on the keyboard while peaking out of the bottom of the headset enabled by small amounts of foam on the headset being cut out.

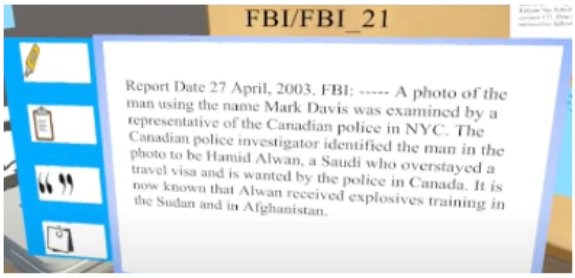


Figure 1.4: Example document provided for analysis with the User-Interface Panel, which is activated on hover on the Left Hand Side. From top to bottom, the buttons are Highlight, Copy to Clipboard, Copy to Clipboard with Citation, and Create Attached Note (Annotation).

Annotations

In IST, annotations are notes that are attached to a document. Meaning when an annotation is made on a document, if the document is later moved, the note will move along with the document. Annotations allow users to offload information directly relating to a document's contents right below it. To create an annotation, participants select the bottom button of the user-interface panel as seen in figure 1.4. Annotations can be hidden at any time by selecting the annotation button again to hide the note.

Note Taking

Similar to annotations, users can create a text-based note using the keyboard within the immersive space. However, unlike annotations, these notes are not attached to a document. These notes allow users to offload information into a more global artifact. To create a new note, participants select the *New Note* button, which is attached to the virtual keyboard table as seen in figure 1.3, right. These notes can be moved around the immersive space freely (i.e., not tied to documents) and placed anywhere within the tracked area.

Label Making

To assist the users in developing organizations within their space, participants have the ability to create labels. Labels are smaller than notes with larger text and are designed for the analyst to be able to specify that documents in a specific region of the space relate to a given topic. Participants select the *New Label* button, which is attached to the virtual keyboard table as seen in figure 1.3 to create a new label. Then, the labels can be moved around the immersive space freely and placed anywhere within the tracked space.

Highlighting

Utilizing the ray coming off from the controller, participants are able to select text within a document. In order to highlight that text, the participant will select the highlight button as seen in figure 1.4. This will then change the selected text from gray to yellow. This interaction works very similarly to selecting text using a mouse/trackpad and then applying a highlighting like in Google Docs or Word.

Searching

Searching in IST, is a global feature, allowing the user to search for a keyword/phrase in the entire document set. To execute a search, the user will select the search button located on

the virtual keyboard table, opening a text-entry box at the location of the tracked keyboard table. Then, the user can type their keyword/phase to enter a search. If that keyword/phrase is found in the document set, the label of the document will turn red, indicating a hit, and the word/phrase is highlighted in the text as well.

Copy/Paste To/From Clipboard

To support the transition between analysis within the immersive space and generating a written report of findings, we developed a feature allowing participants to select text to copy to their clipboard. To copy to the clipboard, a participant first selects text using the ray cast and then selects either the clipboard or quote button as seen in figure 1.4. For example, let's say that a participant is selecting "Mark Davis" from the document pictured in figure 1.4. Once selecting the text and hitting the clipboard button, only the phrase *Mark Davis* will be copied to the clipboard. However, if the participant were to select the quote button, then the phrase *"Mark Davis"* [*FBI/FBI_24*] would get copied to the clipboard. These two features were designed to make offloading relevant pieces of information within IST (i.e., into the search box or into notes) and outside of IST (i.e., into a written report developed on Google Docs/Word) easier.

Save and Load

To support multiple session analysis tasks, IST needed to be able to save the location of all artifacts within the system for loading later. To create a save of all artifacts within IST, a user can select the save button located on the virtual keyboard table as seen in figure 1.3. This save feature would offload all information about an item within the space into a file, including: X_Position, Y_Position, Z_Position, X_Quaterion, Y_Quaterion, Z_Quaterion, W_Quaterion, X_Scale, Y_Scale, Z_Scale, any text input (notes, labels, annotations), and any Highlights applied to the documents. Then, using this save file, we can load the system's state back to where the analysis left off, allowing for multi-session sensemaking analysis tasks to be possible.

1.2.2 Prior work in Immersive Space to Think

Bandyopadhyay et al. developed an initial prototype of IST [5]. This version of IST utilized a tethered Oculus Rift DK2 VR Head Worn Display (HWD). The users had a two-meter by two-meter area in which they were able to move while completing their analysis on the provided dataset. In this version, users were presented with a set of virtual text-based documents that they could grab and move around during their analysis. However, this initial version of IST had no other features to support the user's sensemaking. All of the

documents were displayed to the users in a stack, so users could only view one document at a time until they moved it off the stack.

Using this version of IST, Bandyopadhyay et al. ran a user study (n=8) with the goal of extending Andrews et al.'s study on Space to Think [1] into 3D immersive space to understand if the immersive space provided additional advantages to sensemaking over the 2D space. In this study, they identified two high-level spatial structures formed by the participants - a virtual sphere and virtual curved wall and four different organizational metaphors, including temporal, topic cluster, relevance divide, and working set. Overall, this study found that users utilized the immersive space as a form of external memory and that users could use the space as a form of semantic layer organization, as seen in the prior work in Space to Think. This work suggested many new opportunities for exploration in Immersive Analytics.

Building upon many of Bandyopadhyay et al.'s suggested features and limitations, Lisle et al. developed a new prototype of IST [59]. One important change was that this new version of IST supported a four-meter by eight-meter tracked space where users could physically navigate around while completing an analysis task utilizing a tether-free VIVE Pro with a Wireless Card Adaptor HWD. This new version of IST supported many new features, including single-hand document movement, multi-hand document residing and moving, text scrolling, text highlighting, document duplication, note taking, label creation, keyword search, PDF viewing, and video playback. Note-taking and label creation were achieved by the participant dictating what the label or note should say to the experimenter, and the experimenter would create it for them. Additionally, instead of all of the artifacts being presented to a user in a stack, a virtual bulletin board was used to organize the analysis artifacts so that all artifacts could be viewed without occlusion. Using this updated version of IST, Lisle et al. ran an exploratory user study (n=1) to gather insight into the effectiveness of IST compared to traditional desktop tools for the sensemaking of multimedia datasets. This study suggested that IST was a promising tool for immersive sensemaking and that IST demonstrated potential for the user to organize the information spatially during an analysis task.

To further the findings presented by Lisle et al. in this initial exploratory study with n=1, Lisle et al. ran another exploratory study (n=17) with the goal of understanding how users organize and make sense of a text dataset within IST [60]. In this study, users were asked to complete a historical analysis sensemaking task with the goal of presenting a verbal outline of their analysis at the end of the session. This study identified many interesting patterns within the users' analysis in IST. First, they identified that the users navigated the majority of the tracked space while completing their analysis. Second, they identified three high-level structures that the participants formed during their analysis, including semicircular, environmental, and planar. Lastly, this work identified document selection bias, where the participants interacted with documents in certain locations on the virtual bulletin board more often than other documents.

The initial work on IST has explored how users would complete a text-based sensemaking task within an immersive space. These studies characterized the high-level schemas that were

created during analysis and explored how much of the analysis space the users used while conducting their analysis. This work was foundational in its use of immersive technologies for exploring user behavior during a sensemaking task, which created many future directions for exploring immersive sensemaking.

First, the prior work did not have the users complete a full sensemaking task from foraging up through the presentation stage. In the Pirolli & Card model of sensemaking, the **Presentation** stage is the final step of a sensemaking task. Additionally, the most effort and structure within the process are needed during this sensemaking stage. However, the previous work in IST did not have the user develop a presentation of their findings. Without having the users complete the presentation stage of the sensemaking task, we do not understand how written reports developed during the sensemaking related to the overall organization or structure in the immersive analytic environment. Thus, we don't fully understand how the immersive space would be used during sensemaking.

Second, the prior work only focused on a single sensemaking session. In this exploratory study, the participants were trained on using the system and its features and then given time to complete their sensemaking within the system. While this provided many insights on how the system could be used, it did not allow the users to learn from their prior experiences within the system. Meaning if the participants had more time, they may have organized the documents in the immersive space differently, and the previous work did not capture how this may have changed as the participants became more familiar with the tool. Additionally, it was not very representative of how sensemaking tasks can be completed on multiple days. Thus, we don't fully understand how the organizations used in the spaces may change across an analysis task or how they might change as the participant becomes more familiar with immersive analytics sensemaking.

Third, IST studies have focused on student or novice users, offering valuable insights into how systems like IST can support sensemaking. However, to gain a comprehensive understanding of these systems' benefits and design considerations in supporting the sensemaking task, it is crucial to explore their usage by trained sensemakers. These professionals possess unique critical thinking skills honed through extensive training, making their perspectives invaluable. Consequently, there is a gap in our understanding of how novice analysts compare to professionally trained analysts in the context of immersive sensemaking.

In this work, we want to better understand how IST would be used "in-the-wild" for sensemaking and how we can better support sensemaking through additional organizational tools. Much like the previous work on IST, we want to understand what organizational structures are created and the overall analysis strategies, focusing on understanding how these structures and strategies change over multiple sensemaking sessions. Additionally, we want to understand how novice analysts compare to professionally trained intelligence analysts in immersive analytics sensemaking. Furthermore, we want to better understand how familiarity with the tool affects this system's initial analysis strategies and schemas. Then, utilizing the knowledge gained, we want to understand how additional organization features within

IST affect the sensemaking process, schemas, mental task load, user experience, and task performance during an analysis task. The research questions for this work are as follows:

1.3 Research Questions

RQ 1: How do sensemaking strategies evolve during multi-session analysis in Immersive Space to Think?

Previous work on understanding sensemaking within immersive analytics systems has largely focused on single-session sensemaking tasks. These research studies have provided many preliminary insights into how these systems can be used for analysis tasks. However, these single-session user studies do not evaluate or address how these systems can be used during multi-session analysis tasks where the participant has time to evolve their approach over the analysis task. To address this research question, we conducted two research studies to understand how novice and professional user sensemaking evolves during a multi-session sensemaking task and compare and contrast the similarities and differences between the novice and professional user populations.

RQ 2: How does familiarity affect initial strategies and layouts used for immersive sensemaking?

Along with not evaluating how sensemaking evolves during multi-session analysis tasks, the previous work focused on single-session sensemaking tasks in immersive analytics systems and fails to address how becoming more familiar with an immersive analytic prototype affects the strategies and layouts used during immersive sensemaking. To understand how these tools would be utilized "in-the-wild," we need a better understanding of the overall strategies used and how the organizations developed during these sensemaking tasks change when users become more familiar with the analytic tool usage. To address this research question, we conducted one research study to understand how becoming more familiar with an immersive analytics tool affects the strategies and spatial layouts used during multiple sensemaking tasks. Using the findings from this study, we propose best practices for analysis within immersive space to think, as well as how we can better develop immersive space to think to support these practices.

RQ 3: How can we support best practices for immersive sensemaking with additional organizational features?

Using what we learned from the evolution of sensemaking strategies in immersive space to think from the novice and professional users, along with the insights from understanding how familiarity affects the initial strategies and layouts used, we have developed a much clearer picture of how immersive analytics systems such as Immersive Space to Think can be used for a text-based analysis task. With this new understanding, we explore how to better support sensemaking within immersive space to think with new organizational features. To evaluate

these features, we conducted a user study to understand how these new features affect the overall strategy, spatial structures, cognitive load, and user experience of completing analysis within an immersive space to think.

1.4 Structure of Dissertation

Chapter 1 introduces the background, motivation, and overarching research questions for this work.

Chapter 2 presents a literature review on foundational theories of Human-Computer Interaction, sensemaking, visual analytics, and immersive analytics.

Chapters 3 and 4 address Research Question 1 by presenting two user studies to understand how sensemaking strategies change during multi-session analysis in two different user populations.

Chapter 5 addresses Research Question 2 by presenting a user study evaluating how the 3D structures and sensemaking strategies change with more tool familiarity for immersive sensemaking.

Chapter 6 addresses Research Question 3 by prototyping and implementing new organizational features for Immersive Space to Think to benefit the sensemaking process. Then, these features were evaluated in a controlled user study exploring how the new organizational features affected the sensemaking process.

Chapter 7 concludes with a discussion of the research questions and our findings, as well as detailed future directions of this research.

Chapter 2

Review of Literature

2.1 Foundations in Human-Computer Interaction

The concept of IST draws on two foundational frameworks within Human-Computer Interaction (HCI): embodied/distributed cognition and sensemaking. Before delving into the related work on visual and immersive analytics, we first explore these underlying frameworks.

2.1.1 Embodied and Distributed Cognition

Embodied cognition posits that "cognitive processes are deeply rooted in the body's interactions with the world" [93], meaning that cognition is paired with interacting with the world around us. Wilson proposed 6 claims for embodied cognition, including 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, and 6) Off-line cognition is body based. In the context of IST, two key claims of embodied cognition are particularly relevant: first, the claim that cognition is situated, highlighting the influence of context on cognitive processes, and second, the notion that we offload cognitive work onto the environment. During an analysis task, individuals form schemas of the data, leveraging the environment to organize information in a way that enhances cognitive processes. Complementing embodied cognition, distributed cognition extends the idea that cognition is not confined to an individual's mind but is distributed across individuals, artifacts, and environments [44]. This challenges the traditional notion that cognition is solely an individualistic process, emphasizing the collaborative and distributed nature of cognitive systems.

Together, embodied and distributed cognition underscore the importance of creating meaningful organizations of data during an analysis task to facilitate cognitive processes. Prior work by Robertson et al. [75] provided users with 3D space with landmarks that enabled them to customize their space for organizing information, which increased performance on recalling information within the context of web page retrieval. More existing literature has explored enabling users to organize their cognition within their analysis environment. Works such as Space to Think [1], Jigsaw [86], Maps Around Me [80], Where Should We Put It [66], and previous research in IST [59, 60] have focused on providing opportunities for users to organize their cognition by constructing schemas. Additionally, these tools have aimed to

understand how individuals offload cognitive processes into the environment during sensemaking tasks.

Our work in IST builds upon these foundational ideas to investigate how sensemaking strategies evolve during multi-session analysis tasks, the influence of familiarity on immersive sensemaking strategies, and how additional organizational features can support best practices for immersive sensemaking.

2.1.2 Sensemaking

The Sensemaking Process or Knowledge Generation Process is a complex process that involves taking various streams of unstructured data and forming hypotheses and theories. Pirolli & Card developed a model for intelligence analysis sensemaking as seen in figure 1.1. This model consists of a series of iterative steps which form into two sub-loops the **foraging loop** and the **sensemaking loop** [73].

In the foraging loop, analysts engage in gathering data, akin to assembling pieces in a "Shoebox." This involves the "Search & Filter" phase, where relevant information is identified, followed by the "Read & Extract" stage, where essential information elements are distilled. The loop culminates in "Searching for Information," leading to the creation of an "Evidence File" comprising snippets or dots of low-information density that can be instrumental in subsequent sensemaking. This phase emphasizes the iterative nature of sensemaking, allowing analysts to navigate up and down the process stages as needed.

Transitioning to the sensemaking loop, analysts leverage the accumulated evidence to form schemas and hypotheses. This stage involves activities such as "Schematizing," "Building Cases," and "Searching for Support/Relations." The culmination of this loop is a presentation stage, where findings are formalized into intelligence briefings or written reports.

The process is bi-directional and iterative, where analysts work up and down the process stages as needed while completing sensemaking. For example, after developing an initial hypothesis, analysts may need to go back to steps to gather more information from external data sources or the shoebox to gather support for their hypothesis. Going beyond just modeling the sensemaking task, Klein et al. present how the sensemaking task requires creativity, comprehension, mental modeling, and situational awareness [52]. We formulate mental and physical models to organize the "unknown" to ease the cognitive load during the sensemaking task. These structures or frameworks allow us to better understand the underlying data collected for analysis [90].

However, one criticism of the Pirolli & Card model is that it can be perceived as too linear and focused on information processing, potentially lacking a direct representation of the intelligence analysis process itself [47]. In response to this, Klein et al. propose the *Data/Frame Sensemaking Process*, which introduces a less linear model as seen in figure 2.1. The model contains two distinct loops: a backward-looking loop of "Mental Model Forma-

tion," where existing information modifies frames, and a forward-looking loop of "Mental Stimulation," where frames are adjusted through anticipation. This model introduces the concept of "frames" as metaphors for perspectives or viewpoints shaped by the data. Analysts can have a hypothesis of an event, and as they gather new data, they elaborate on the frame, adding new data into the frame and inferring new relationships. Contradictory information may lead to questioning and re-framing of the overall hypothesis. This model allows for continuous data modification through re-framing, preservation, and elaboration.

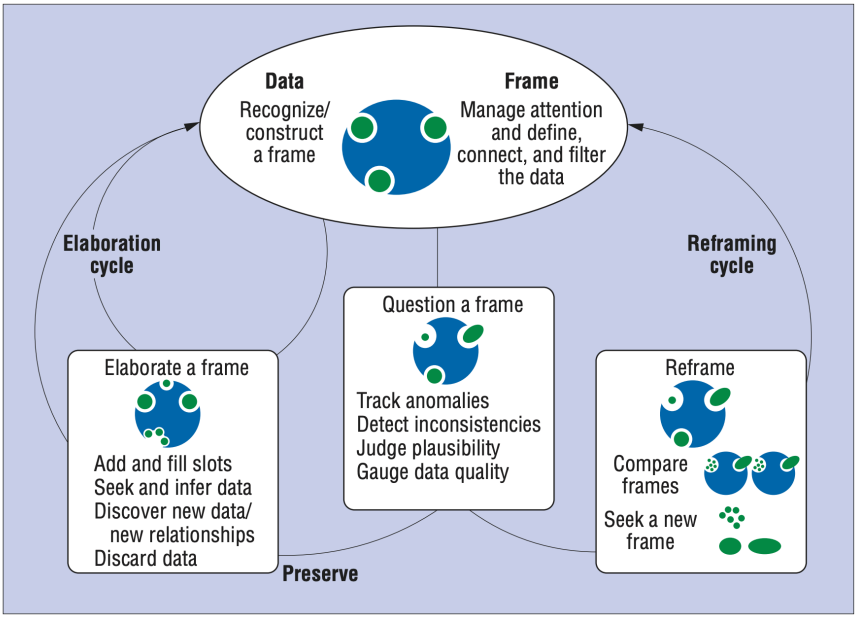


Figure 2.1: Klein & Moon Data Frame Sensemaking Model. This model focuses on creating mental model frames of the data where these frames are constantly refined, adjusted, and elaborated.

In the realm of sensemaking, regardless of the specific model in consideration, the overarching process involves transforming largely unstructured raw data into a hypothesis, searching for supporting evidence, and presenting refined findings. While this fundamental process may seem straightforward on paper, its complexity amplifies when applied to professions where sensemaking is at the core, such as intelligence analysis. The cognitive demand inherent in completing sensemaking emphasizes the crucial role of supportive technologies in this domain. Visual analytics stands out as a field that concentrates on providing such support through interactive visual interfaces [89].

These sensemaking models effectively capture the intricacies of the intelligence analysis process. In our work in IST, we delve into the evolution of sensemaking strategies over multi-session analyses, recognizing that sensemaking often unfolds across multiple sessions. This differs from traditional user-study designs, which primarily focus on single-session tasks and align with the reality of sensemaking's extended duration. Through our approach, we aspire

to comprehensively evaluate how IST supports sensemaking across these prolonged durations, contributing valuable insights into how these tools could be used "in the wild."

2.2 Visual Analytics

To address the challenges in the sensemaking process, visual analytics focuses on providing tools to assist users in analysis tasks. This section defines visual analytics, presents a conceptual model, and introduces examples of visual analytic tools.

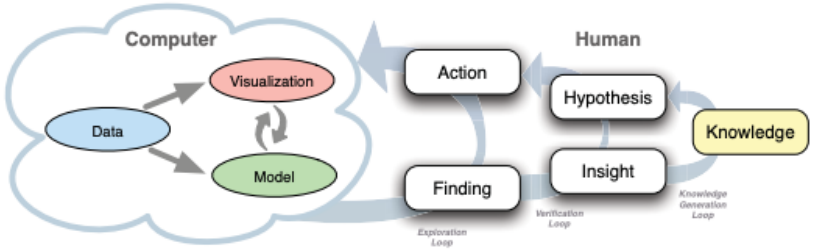


Figure 2.2: Sacha et al.'s Knowledge Generation Model for Visual Analytics. The model consists of two parts: the computer and the human. In this model, the computer or left-hand side illustrates the visual analytic system and the right-hand side indicates the knowledge generation process. The right-hand side has three sub-loops - exploration, verification, and knowledge generation. This model highlights the human-machine teaming needed in visual analytic systems.

2.2.1 Definition & Model

Thomas & Cook defined Visual Analytics (VA) as the "science of analytical reasoning facilitated by interactive visual interfaces" [89]. Keim et al. presented 4 main goals of visual analytic systems: being able to synthesize information, detect expected/unexpected information, provide timely support, and present insights in an actionable manner [49]. Additionally, they highlight that VA is more than just data visualization; it is an intersection between decision-making, visualization, and data analysis. However, as Sacha et al. point out, the model for VA presented by Keim et al. [49, 50] focuses more on the system and less on the human. Building upon the sensemaking process and focusing on including the human process within the model, Sacha et al. presented a model for the **knowledge generation process for visual analytic tools**, which can be seen in figure 2.2. This model consists of two key components, the *human* and the *computer*. In this model, the computer component represents the visualizations, analytics, and data, while the human component represents the cognitive processes undergone during a sensemaking or analytic task. A key aspect of

VA is that the analyst is at the center of the process, guiding the analytics and visualizations created by the computer. The goal of the computer, on the other hand, is to aid the analyst in synthesizing the large amounts of data presented through these interactive methods.

VA places the analyst at the center, guiding computer-generated visualizations and analytics to aid in synthesizing large datasets. However, as Endert et al. pointed out, there is a lack of support for the entire sensemaking process from foraging to presentation [27]. This highlights that most VA tools focus either on assisting in the foraging loop or the sensemaking loop but not on the entire process together. In this work, we explore how IST can support the sensemaking process as a whole in an attempt to design tools that can be used during the whole knowledge-generation process.

2.2.2 Visual Analytics Examples

Various tools have been developed in the field of VA to assist in sensemaking tasks. Here are examples of VA tools that support users in analyzing non-quantitative datasets, as our work builds on non-quantitative datasets with text-based datasets.

Jigsaw: Coordinated Views for Document Analysis

One highly influential tool for text-based VA is Jigsaw [36, 37, 38, 86]. Jigsaw is a tool that provides multiple coordinated views of a document set through visual connections on shared entities. The system's main focus is to provide connections between entities in the documents. This system has 4 different views or ways to visualize the documents and entities, including 1) tabular connections, 2) semantic graph, 3) scatter plot views, and 4) the text document itself. In the tabular connection view or list view, entities are connected via lines and color highlighting is used to show a stronger connection. In the graph view, entities are extracted from the text documents, and edges are drawn between the reports and entities. In the scatter plot view, entities are shown on the x and y axis, and the dots are reports or text-documents which contain both entities. Lastly, the text document is the text with highlighting and underlining based on entity extractions. Within Jigsaw, these different visualizations or views are synchronized, where interaction with one view will translate to all of the other views. In a reflection paper on the Jigsaw system Görg et al., present some key lessons learned on the Jigsaw system, including but not limited to the following: 1) "Reading the documents still matters" where it is used to help build and expand knowledge about the plot and knowledge of the dataset as a whole. 2) "Do not neglect evidence marshaling", where they discuss the important of offloading hypotheses and findings to keep track of the analysis as an essential feature, especially with tasks that are carried out over a longer period of time (i.e. not single-session).

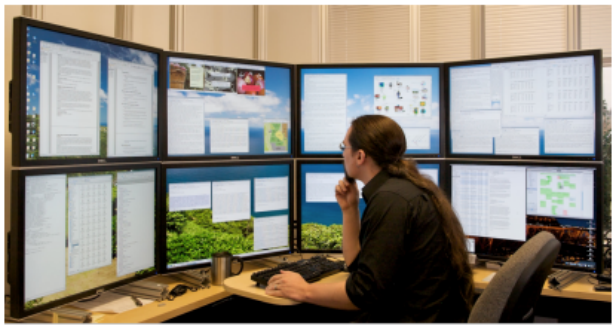


Figure 2.3: The **Space to Think** or **Analyst Workstation** presented by Andrews et al. It was made up of an array of 4 x 2 30" monitors that were slightly curved around the user [1].

Space to Think: Expanding the Analysis Environment

In contrast, Space to Think [1] takes a unique approach to VA by providing a large physical space to use while completing a text-based analysis task. In this work, they created an *analysts workstation*, which consisted of an array of 4 x 2 30" monitors combined together to create a large high-resolution analysis space as seen in figure 2.3. This work aimed to understand how the space afforded by the *space to think* supported sensemaking. In this work, they found that providing the users with a large space for an analysis task supported their sensemaking process by allowing them to externalize cognition into the environment around them. For instance, they found that participants spread documents out across the monitors building organizations, including "work zones", "timelines," and "clusters" within the workspace. Additionally, they found that this externalized cognition was accessed via physical navigation, rotating the chair, to retrieve information from the environment. However, Space to Think is still limited to 2D organization and traditional desktop interactions via mouse and keyboard.

Force/Star SPIRE: Leveraging Semantic Interaction

ForceSPIRE and StarSPIRE [12, 25] share a common emphasis on the spatialization of documents during analysis, akin to the approach seen in Space to Think. However, these tools go further by integrating semantic interaction to provide enhanced assistance to the user. ForceSPIRE and StarSPIRE are tools that leverage semantic interaction to provide a user with assistance using a sensemaking task. In both systems, the user's interaction with the text-based documents steers the underlying model. Additionally, StarSPIRE, as presented by Bradel et al. [12], focused on creating multi-model semantic interaction to allow for use with both small and large data analysis tasks. In ForceSPIRE and StarSPIRE, the visualization presented to the user is a node-link diagram that uses a modified force-directed layout. In these systems, documents are represented as nodes, and edges represent shared entities

between documents. During analysis within these systems, users can execute the following types of interactions: moving/pinning nodes, resizing a document, minimizing/closing a document, annotations, searching, highlighting, and overlapping documents, which then influence the underlying model parameters. Both of these approaches focus on the idea of specialization as a key aspect in the analysis process, and additionally, ForceSPIRE was designed on the concept of Space to Think as well as StarSPIRE being able to scale to any size display.

As technology is ever evolving, the field of VA has seen many advancements, and its work has paved the way for a new research area called **immersive analytics**. With the emergence of lower-cost immersive technologies, evaluating how these technologies can enhance sensemaking is becoming increasingly important. The transition from traditional VA tools to immersive analytics introduces opportunities for more immersive and interactive data exploration, offering new perspectives and enhancing the sensemaking process.

2.3 Immersive Analytics

Expanding on the principles of VA, the field of immersive analytics explores the use of more immersive technologies to enhance sensemaking tasks. This section provides an understanding of immersive technologies, a definition of immersive analytics, and examples of relevant immersive analytics systems.

2.3.1 Immersive Technologies

Immersive technologies offer various ways to go beyond traditional workstations, providing enhanced sensory input/output channels, increased fidelity, new user input methods, and more natural interactions [84]. For example, Skarbez et al. discuss how support for additional channels of sensory input (e.g., olfactory or haptics), the capability to render sensory output at higher fidelity (e.g., increased field of view/regard, increased display resolution, or refresh rate), support for new types of user input (e.g., voice or haptic input, head-/hand/eye/body tracking), and support for more natural interaction (e.g., physical navigation, hand tracked interactions) can all lead to more immersive technology than traditional workstations [84]. Some examples of immersive analytics technologies that support these more immersive characteristics include Large High-Resolution Displays, Virtual/Augmented Reality, CAVEs (Cave Automatic Virtual Environment)[18, 31] and Tabletops/Workbenches [54].

The benefits of immersion, such as improved depth cues, reduced clutter, enhanced spatial orientation, and increased engagement [2, 10, 68], open new possibilities for sensemaking

tasks. Our focus is on Virtual/Augmented Reality displays, particularly headsets evolving from Sutherland's pioneering head-worn display [87]. These headsets provide stereoscopic 3D visuals, offering a sense of depth and presence within virtual environments. Virtual Reality immerses users entirely in computer-generated worlds, while Augmented Reality seamlessly integrates computer-generated elements into the real world.

2.3.2 Immersive Analytics Defined

Immersive Analytics (IA) builds upon VA concepts. According to Chandler et al., the overarching objective of IA is to "explore the applicability and development of emerging user-interface technologies for creating more engaging and immersive experiences and seamless workflows for data analysis applications" [14]. Skarbez et al. later defined IA as "the science of analytic reasoning facilitated through immersive human-computer interfaces," emphasizing the importance of the sensemaking process [84]. These interfaces are ideally designed to aid in knowledge generation through abstract data visualizations and offer the advantages of embodied interaction [22, 84]. Additionally, they contribute to "data understanding and decision-making" [68]. Unlike VA systems, IA systems introduce novel possibilities by providing enhanced depth cues, an opportunity for 360-degree 6 degrees of freedom interaction space, minimized distractions, and improved spatial comprehension [1, 10, 68]. These immersive technologies enable users to metaphorically "step through the glass" of traditional display desktop technologies, placing them inside the data [85].

Early IA studies show that these systems can aid in externalizing memory [1] and recall [53]. The field of IA is rapidly evolving, with numerous open challenges [29]. Key focus areas of IA include Situated Analytics/Spatial Data Analysis, Collaboration, and Immersive Data Visualization.

Situated Analytics / Spatial Data Analysis

Situated Analytics refers to an approach in IA where the data itself is represented within the context of the environment where that data is generated, collected, or utilized [91], with Embedded Data Representation taking this definition a step further focusing on how the data must overlay the physical referent (the physical entity to which the data is referring) [92]. In situated analytics, the data and environment are presented to the user in a way that provides an increased understanding of the spatial context in the analysis, where providing the data in situ can enhance the decision-making process. Some examples of situated analytics include Elsayed et al.'s. work [24], which focused on providing overlays on the product in a grocery store to assist in searching, filtering, and ranking products. Another example of situated analytics includes Rowden et al.'s work on visualizing Wi-Fi signals in a 3D model of a building. This work focused on exploring different visualization techniques for localization, strength, and coverage [77, 78].

Spatial Data Analysis in IA focuses on understanding relationships, patterns, and trends in data with a spatial or geographical component. Examples include Satriadi et al.'s *Maps Around Me* [80] and Yang et al.'s work on Choropleth Maps [96].

Collaborative Immersive Analytics

Another large topic area of IA is Collaboration or Collaborative Immersive Analytics (CIA), which focuses on utilizing immersive technology to provide a novel way to collaborate with others during a sensemaking or data analysis-driven task. Billinghurst [8] made use of the time-space taxonomy: **time** (synchronized / asynchronous) and **space** (Co-Located / Distributed) from Computer-Supported Collaborative Work (CSCW) [4] within CIA, with Sereno et al. [82] furthering this taxonomy to include **role symmetry**(symmetric/asymmetric) and **technology symmetry** (symmetric/asymmetric). Some examples of CIA include Reipschläger et al. [74], who utilized large interactive displays to create a collaborative environment while still including personalized views using augmented reality headsets, where the collaborators HWD were able to generate individual views in the headsets while completing the analysis. Another CIA system with a mix of immersive technologies is Uplift [28]. In Uplift, a combination of tabletop, tangible widgets, and augmented reality was used to create a "walk-up and use" collaborative system co-designed with local stakeholders. Cordiel et al. [16] also evaluated differences between CAVE-style and HWD for collaborative analysis in a network connectivity task. While our work is not directly focused on CIA, a future direction of IST is to understand how users would use a system like IST for a collaborative sensemaking task.

Immersive Visualization

While all of the examples we provided above are focused on data visualization, we have separated out the above examples into more detailed categories such as situated analytics and collaboration. However, immersive analytics research opens up new opportunities for 3D immersive data visualization, which takes advantage of the additional dimension for visualizing the data used in these analysis tasks. For example, a collaborative data visualization system from Lee et al. was Free-roaming Immersive Environment to Support Team Analysis (FIESTA) [56, 57]. In FIESTA, collaborators wear wireless virtual reality HWDs, which allow the users to freely walk around in the immersive space while completing their analysis with quantitative data, where the collaborators can generate visualizations simultaneously within the system to gain insights into the data. Another example of a more focused immersive data visualization system is by Cordeil et al. with ImAxes (Immersive Axes), which leverages embodied interaction for creating visualizations within IA systems [17].

2.3.3 Large Workspaces in Immersive Analytics

Our work is based on the idea of providing large workspaces for completing sensemaking tasks. In our work, we aim to understand how users complete sensemaking within these systems, what structures users form while completing analysis, and how we can better assist users in organizing their spaces. There is a lot of research surrounding large workspaces in immersive analytics, including understanding how the users utilize the space [59, 60] to evaluating the cognitive load of completing analysis of these systems [34].

One deeply researched area has been in understanding the deeper organizational structures made within the spaces while completing an analysis task [19, 20, 65, 66, 80]. Our work aims to understand how the structures made change during multi-session sensemaking tasks, as well as understand how familiarity with IA analysis affects the layouts made. Additionally, Satriadi et al.'s work also focused on understanding how users organize information in a map analysis task [80]. In this work, they explored the overall structures that the users formed while completing a series of tasks, including a naïve layout, search, route planning, comparison, and general layout tasks. In this study, they identified patterns in the type of layouts that the participants formed, including planar, spherical cap, and spherical, as well as patterns related to overview-detail relationships such as central window, occluding, and coordination.

Another work that aimed to understand where users organize information during an analysis task in large workspaces was "where should we put it?" [66]. This work aimed to understand how users organize information in an AR environment with a fully-furnished and side-furnished (some furniture existed but was pushed up against the walls) condition. This explored how both individual and collaborative users used the furniture as a way to scaffold information and to see if the existence of furniture affected the overall organizations in their environment. This work identified 8 different layout strategies including grid-like, cylindrical, furniture-anchored cylindrical, furniture-aligned cylindrical, vertical surface furniture-based, omni-furniture-based, horizontal surface furniture-based, and canvas centered. This work found that the users actively used the furniture for organizing the information during a sensemaking task.

Another area of researching large workspaces by Lisle et al. aimed to understand how much space is beneficial for a sensemaking or analysis task [62]. In this work, they evaluated a small (virtual monitor), large (multi-virtual monitor), and immersive space for use during a text-based sensemaking task. This work found that immersive and large spaces provided greater satisfaction and decreased frustration when compared to the small single monitor condition and that users utilized more spatial memory as the amount of space provided for analysis increased. Another work, by Lisle et al., aimed to understand which technology (AR or VR) was better for immersive analytics in a text-based sensemaking task. This work found that both technologies afforded different benefits, with VR providing increased focus and AR allowing users to access tools from the environment around them (e.g., whiteboards), suggesting that future work for IA sensemaking can leverage benefits from both AR/VR

by providing a tool where users can decide where on the XR continuum to complete their analysis.

Chapter 3

Exploring the Evolution of Sensemaking in Immersive Space to Think

Previous studies in IST have demonstrated advantages for sensemaking, yet they primarily concentrated on single-session tasks, leaving the entire sensemaking process unexplored. An unexplored area lies in investigating how users adapt their analytical and organizational strategies when undertaking an entire sensemaking task within an IA system forming Research Question 1 (RQ1): How do sensemaking strategies evolve during multi-session analysis in Immersive Space to Think? To address this question, we conducted a study involving novice users to gain insights into how these systems are used throughout the entirety of a sensemaking task.

3.1 Introduction

We are living in a world of Big Data [79]. With the vast influx of data, we need effective technologies to help us parse through a large amount of data and make sense of it. With the growth of big data and its importance over the past ten years [79] a simultaneous acknowledgment of the value of information visualization has taken place. Aiming to build views of this big data, work in information visualization allows analysts to make more informed hypotheses and decisions through better representations of trends, clusters, and outliers in their data.

In recent years we have seen the advent of low-cost Virtual Reality/Augmented Reality (VR/AR) display technologies. As these technologies are becoming widely available in consumer markets, we have seen a growth in the field of immersive analytics, which aims to investigate how immersive technologies can be used to support analysis and decision making. Immersive analytics was first introduced by Chandler et al. [14] in 2015. In 2019, Skarbez et al. [84] redefined immersive analytics as "the science of analytical reasoning facilitated by immersive human-computer interfaces" to provide a more flexible definition. This definition focuses on analytic reasoning, which refers to the computer-aided sensemaking process (knowledge generation)[73]. The sensemaking process consists of two loops: information for-

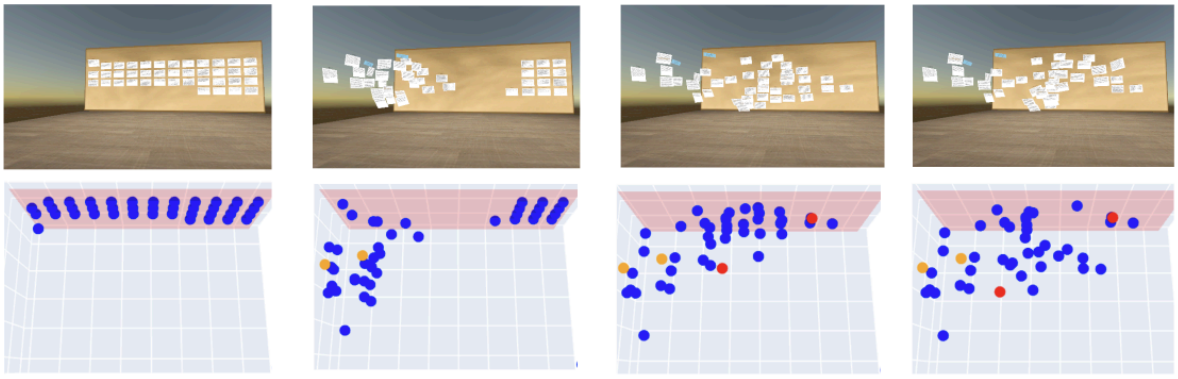


Figure 3.1: Overview of an Immersive Space to Think Analysis Space. This example is of participant five's spatial layout over time from left to right: start, end of session 1, end of session 2, and end of session 3. Top Row: view of immersive space. Bottom Row: view of 3D scatterplots used for analysis in a top-down orientation.

aging and the sensemaking loop. The goal of immersive analytics is to unobtrusively support both parts of these processes in a way that aids the analyst's cognition.

One approach to immersive analytics is exemplified by the ImAxes system [6, 17], which focuses on immersive data visualization. Another unique approach that focuses on text-based data includes Space to Think [1] and Immersive Space to Think [60], which provide a vast amount of space for structuring the unknown by offering opportunities for distributed cognition [44]. These immersive analytic systems offer a way to externalize your memory into the environment. However, one limitation of the existing research is that it focuses on a single analysis session. Moreover, in an application such as intelligence analysis, analysis can be ongoing for months or years at a time [47]. In our work, we focus on a sensemaking task completed across three sessions to simulate a long-term sensemaking task that leverages the ability to externalize cognition into the environment. In this chapter, we build upon an existing immersive analytic software prototype - Immersive Space to Think (IST) [60]. Using this software prototype, we conducted an exploratory three-session user study with the following goals: 1) Understand how immersive technologies affect the overall sensemaking task from foraging to presentation, 2) Understand users' 3D organizational structures and how they evolve during a temporally distributed sensemaking task, 3) Understand how immersive technologies can be used in combination with more traditional computing technologies while completing the sensemaking task, and 4) Explore ways to analyze a distributed sensemaking task that could be used for future immersive analytic sensemaking studies.

3.2 Immersive Space to Think Prototype

3.2.1 New Interaction Techniques

Document Randomization In our prior work, we found participants had a strong document selection order bias [60]. To mitigate the effects of document selection order bias, we developed a randomized loading feature. When documents are loaded into IST, all documents are shuffled before placement onto the bulletin board. By providing this feature during our user study we hoped to prevent the effects of ordering bias.

Text Entry To address the limitations of the previous wizard-of-oz technique, we developed a new text-entry feature. Using a tracked physical rolling table and keyboard, as seen in figure 3.2, participants can enter text with a standard device and can move the keyboard around the immersive environment with them. The table and keyboard are represented with similarly sized 3D models in the virtual environment. However, the user's hands are not tracked. So users must place their hands properly on the keyboard using the sense of touch or by viewing their physical hands and keyboard outside the headset.

Copy/Paste to/from Clipboard To support the later stages of the sensemaking process, we wanted to provide the analyst with a way to export information from IST into traditional workspaces. We provide the users with a copy and paste button to get text from the immersive space onto the computer clipboard, acting as a hybrid user-interface [32]. From there, the users can paste the information in the immersive space or into other applications.

3.3 User Study

3.3.1 Goals and Research Questions

In our study, the primary goal was to understand how IST supports the complete sensemaking process, from foraging to report generation, over a longer period of time (multiple sessions). Additionally, we were trying to understand the relationship between the IST spatial structure and the report's structure. Lastly, we aimed to explore how users blend the use of traditional workspaces (laptops, keyboards, text-editors, etc.) with 3D immersive spaces. With these goals in mind, we generated the following research questions:

Spatial Structures

- ***RQ 1.1A*** What spatial structures do analysts form in IST, and how do these spatial structures change over multiple sessions?
- ***RQ 1.1B*** How do the 3D spatial structures map to the 1-dimensional, hierarchical reports?

Physical Navigation

- ***RQ 1.2*** How do analysts navigate the virtual space while conducting analysis and writing the report?

Bridging Technologies

- ***RQ 1.3*** How do analysts use both traditional and immersive technologies simultaneously during the presentation stage of the sensemaking process?

Interactions/Strategies

- ***RQ 1.4*** How do the interactions/strategies of participants correlate to the quality of the reports generated?

These questions seek to understand how immersive analytics can be used to support the entire sensemaking process (foraging to presentation) while completing a dynamic sensemaking task. Additionally, these questions aim to address how we can better support the most time-consuming stage of the sensemaking process – Presentation – as quantified by Pirolli and Card [73].

3.3.2 Apparatus

In our experiment, we used a wireless HTC VIVE Pro Head Worn Display running on a desktop PC. The desktop PC had an Intel Core i7-8700K processor and an NVIDIA GeForce GTX 1070 Ti graphics card. The user held one Valve Index wireless controller in their dominant hand to interact with the documents. The 3D interaction technique used in this IST prototype was Ray-Casting with reeling. The document size was 0.5 x 0.3 (m), which supports about one paragraph of text (Text-Mesh Pro text size at 40) without the need for scrolling. For text input, participants used a wheeled table with a wireless keyboard, as seen in figure 3.2. To provide better support for text entry, small amounts of foam were cut off the HTC VIVE Pro headset to allow users to look down and see their fingers on the physical keyboard. Additionally, for writing both the outline and report, participants used a 2019 Mac Book Pro with Chrome Remote Desktop connected to the desktop PC. A SteamVR 2.0 Lighthouse tracking system tracked user movements, which covered a four-by-eight meter space.

3.3.3 Experimental Task

This experiment aimed to understand how immersive analytic systems can support the dynamic nature of the entire sensemaking process, from foraging to presentation. For the task,

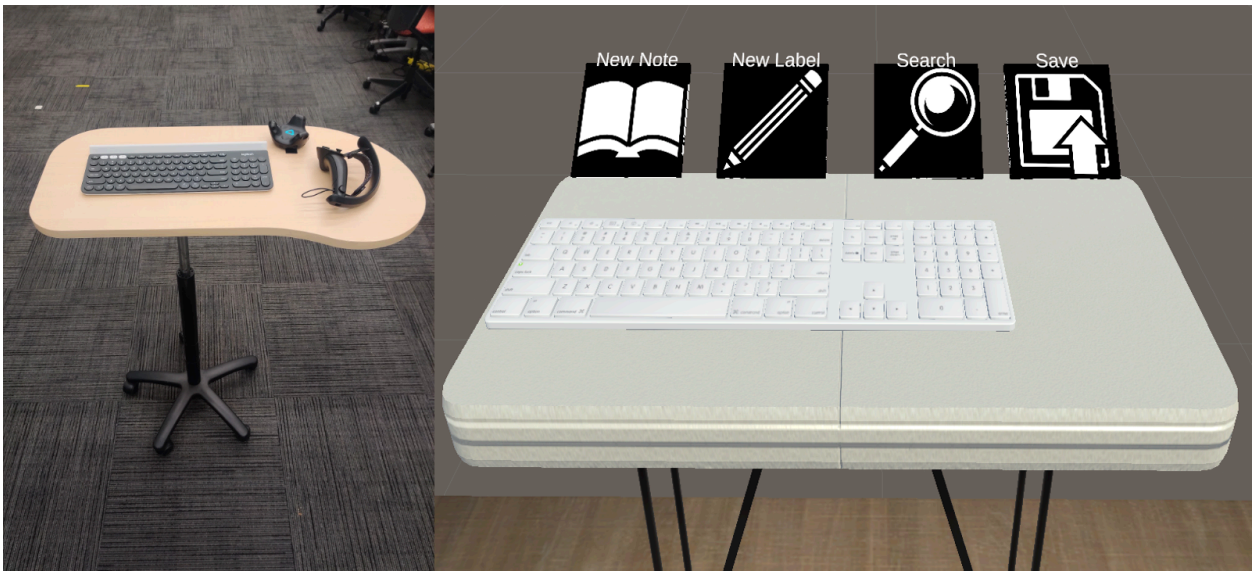


Figure 3.2: Left: Keyboard stand with wireless keyboard, HTC Vive Tracker, and example space for a controller. The keyboard is on a wheeled table which allows the participant to move the keyboard around the space as needed during their analysis. Right: Virtual representation of the keyboard. Attached to the keyboard are four menu options. These options from left to right are *New Note, New Label, Search, and Save*. These menu options are attached to the table and will move with the table as it is moved around the space.

participants were asked to analyze a set of text-based documents, of which the participants would have no prior knowledge, with the goal of writing a report of their findings at the end. The task provided to the participants was as follows:

Today is April 27th, 2003. You are an intelligence analyst working for the federal government. It is believed that terrorists are planning an imminent attack on the United States. Other analysts have gathered a set of potentially relevant documents containing information about potential suspects. These documents have been loaded into the Immersive Space to Think system for your analysis.

Your goal is to analyze the information and develop a specific hypothesis about any potential planned terrorist attack(s) against the US. Your hypothesis should identify ***who, what, when, and where****.*

Over the course of 3 sessions, you will develop a report for the Office of Homeland Security. To do this effectively, you must prepare a defensible and persuasive report that describes exactly what your conclusions are based on the documents provided. During session one, you will focus on analyzing the documents and preparing a specific hypothesis of your findings (who, what, when, and where). During session two, you will continue your analysis and prepare a written outline of your findings. During session three, you will finalize your analysis and write a 1-2 page report of your findings. Your report must state what action or

actions the terrorist(s) are planning, where they will occur, and when they will occur using evidence and citations from the documents provided. Our hope is that your report can be used to thwart the terrorist(s) threat.

Participants were informed that both the outline and the final report were checkpoints that they were expected to have completed by the end of sessions two and three, respectively. However, they were also advised that they were allowed to work ahead or return to tasks as needed to support their own sensemaking process. In the analysis of the participants' strategies, we saw participants work ahead in completing the checkpoints as well as returning to analysis as needed.

3.3.4 The Dataset

The goal for the dataset was to be complex in its features, making it challenging for analysis. Ideally, this dataset should have many different features that force the participant to use the space around them to create organizational structures that are meaningful, similar to the organization structures seen in Space to Think [1] and previous iterations of IST [60].

The dataset we used for this study is a fictional intelligence analysis dataset called **Sign of The Crescent**, which contained a total of 40 documents. Each document is about a paragraph in length. The features of the dataset include aliases, addresses, phone numbers, bank account information, places of business, etc. Additionally, this dataset contains both relevant documents (23) and distractor documents (17). Along with relevant/irrelevant documents, there are three major plotlines contained in this dataset connected through documents called coordination documents. Plotlines are defined, in this dataset, by the location in which the events are happening (i.e., Atlanta, Boston, or New York). Plotline 1 contains five documents, Plotline 2 contains seven documents, Plotline 3 contains four documents, and there are seven coordination documents.

Along with all of the features described above, the documents purported to come from three different reporting organizations: FBI, CIA, and Sanctioned Intercepts. There were 24 FBI documents, 11 CIA documents, and 5 Sanctioned Intercept documents.

3.3.5 Participants

We recruited eight university students for our study, including two females, five males, and one non-binary, with a mean age of 22.6 (standard deviation of 3.85). Five of the participants had tried VR or AR technologies once or twice; one participant had tried VR or AR 3-10 times, and two participants had tried VR or AR technologies more than 10 times. All participants had normal or corrected vision. The participants were recruited using internal university email lists. This study was approved by Virginia Tech's institutional review board.

Participants were compensated for their time in this study pro-rated by how many sessions they completed. Session 1 – $20.00, Session 2 – $20.00, and Session 3 – $30.00. The payment was made in cash at the end of the experiment.

3.3.6 Measures

Before the study, participants were asked to take a pre-study questionnaire that asked them general background information such as gender, age, VR/AR usage, etc. A log file was generated during the study that tracked the participant's head, controller, and keyboard position and orientation. This data was reported at a rate of about 10 logs/second. Additionally, this log file tracked every action the participant conducted within the immersive space, e.g., highlighting, moving a document, or searching.

In addition to the log file, a record of the spatial layout of all documents was generated once a minute for the entirety of the user study. These save files created a minute-by-minute snapshot of the participant's analysis in the immersive space.

A screen recording was taken from the first-person point of view of the participant conducting the analysis. After each session, the participants were interviewed about session-specific questions, and after the study, the participants took part in a post-study semi-structured interview.

Lastly, we also collected both an outline (generated during session 2) and a report (generated during session 3) of the participant's findings from their analysis. These documents were created using Google docs with the laptop placed on the tracked table.

In addition to the data collected for this study, we worked with a set of professional intelligence analysts to develop a grading rubric for the final reports. The rubric was designed to produce subjective quality scores and objective correctness scores. The quality score measured conciseness, persuasiveness, clarity, completeness, relevance, and bias—the correctness score measured plotline level details (Who, What, When, Where). A copy of the report grading rubrics can be found in the appendix.

3.3.7 Procedure

The user study was broken down into three 75-minute sessions. Participants were encouraged to schedule all 3 sessions in the same week and if possible on 3 separate days. Each session was broken down into different stages: pre-session, tutorial, session, and interview. Details on each of the sessions can be found below.

Session 1

Pre-Session Before arrival each participant was screened for our COVID mitigation protocol. Upon arrival, participants were asked to review the informed consent form and ask questions if they had any. After signing the form, participants were asked to fill out a questionnaire to answer some demographic and background questions.

Tutorial Participants were trained on how to use both the HTC VIVE Pro headset and the Valve Index controller. After the training, participants would don the headset and begin the IST tutorial. Using a practice dataset, the experimenter taught the participants the immersive boundary limits for walking, how to manipulate documents, how to use the wheeled keyboard/table, and all of the features related to the IST system. After reviewing all of the features, participants were given 5 minutes to practice using the features, walking with the headset, and typing using the keyboard on their own.

After finishing the tutorial phase, participants were asked to take off the headset and have a seat while the experimenter loaded up the experimental dataset. During this time, the participants were given a printout of the experimental task (section 3.3.3). Participants were encouraged to ask questions if clarification was needed and when ready, participants were asked to don the headset once more to begin their analysis.

Session Before beginning, participants were instructed to start their analysis and develop some initial hypotheses. Participants were given about 40 minutes for this task. Participants were encouraged to ask questions on the task or the system if they needed help.

Interview Session 1 ended with a 5-minute post-session interview. This interview phase was designed to gather information about their overall strategy for analysis and overall spatial structures/organizational pattern that they had formed.

Session 2

Pre-Session Session 2 began with a quick pre-session questionnaire which asked about tiredness and how they were feeling.

Session The goal of this session was to continue with analysis and develop an outline of their findings. The participants had about 65 minutes for analysis and outline writing.

During the outline stage, the participants used a laptop on the wheeled table stand to write their outline with Google docs. In order for the participant to see the laptop, participants were informed they could either lift the headset lenses off their eyes and rest the headset on the top of their head or continue to peek out of the bottom of the headset. The participants were encouraged to use the immersive space as much as needed while developing the outline and were able to use *copy and paste* features to get text from the immersive space and paste it into their working outline. A 15 minute warning was provided as a reminder for participants

to work on their outline however, the participant could begin the outlining stage early using IST artifacts such as notes or annotations.

Interview This session ended with a post-session interview which aimed to answer questions on strategy, spatial structure/organizational patterns. Additionally, this interview asked questions about the transformation process used to move from the spatial organization of documents and notes to the outline.

Session 3

Pre-Session Session 3 began with the same pre-session questionnaire as session 2. Following the questionnaire, participants were reminded that they would be writing the report of their findings during the session.

Session The goal of the last session was for participants to develop a final report of their findings. Participants were given about one hour to complete their analysis and to write the report of their findings. During this time, the participants were able to use the headset to see their immersive space, and use the tracked laptop and table to write their report in a Google Doc.

Interview After completing the report, participants took part in a final 15-minute semi-structured interview. During this interview they answered questions about the task including: overall strategy, overall spatial layout, and how the layouts changed over time. Additionally, the participants answered questions about they system as a whole and what additional features would be helpful for analysis.

During each session, participants were given a time warning at the halfway mark and given the option to take a break if needed. Additionally, time calls were given at 15 minutes, 5 minutes, and 1 minute before the end of the session.

3.4 Results and Discussion

To address RQ1.1, we looked at the overall spatial structures formed by the participants, how the spatial structures changed over time, the deeper organization patterns used during analysis, and mapping of the spatial layout to the report structure. To address RQ1.2, we examined how the user moved through the space and how much time they spent in different parts of the space across sessions. To address RQ1.3, we looked at how users bridged multiple technologies. Finally, to address RQ1.4, we investigated report scores and the overall strategies of the participants.

3.4.1 RQ1.1A: Spatial Structures and Change Over Time

Spatial Layouts

Using a bird's-eye view of each participant's immersive space after the task was complete, we can see the overall organization pattern each one used for analysis. We examined the overall placement of the documents, notes, and labels to look for patterns within the data. In previous work conducted by Lisle et al. [60], there were three overall spatial layouts formed by participants: Semi-circular, Environmental, and Planar. Using these organizational structures, we categorized each of the participants' final spatial layouts. In our study, we had seven semi-circular/circular layouts and one environmental layout (P5). Each of the final spatial layouts can be seen in figure 3.3.

Change Over Time

In the development of the experimental task, we aimed to target different stages of the sensemaking task in each session with some flexibility given to the participants as needed. Session 1 focused on the foraging and evidence file stages of the sensemaking task. Session 2 continued foraging and began the schema formation stage. Lastly, session 3 focused on the presentation stage of the sensemaking process. Diving deeper into RQ1.1A, we wanted to understand how the organization patterns (spatial structures) changed over the course of the temporally distributed sensemaking task. In order to evaluate change over time, we looked at overall document movement patterns as well as overall organizational patterns.

We first examined the number of documents moved in each session, as seen in figure 3.4. We have a median of 28 documents moved during session 1, 31.5 during session 2, and 11.5 during session 3. Additionally, we see a statistically significant difference between sessions 1 and 3 ($p = 0.04$) and sessions 2 and 3 ($p < 0.001$) using a T-Test for Independent Samples with Bonferroni's Correction.

We also looked at the distance that those documents moved, as seen in figure 3.4. The median movement distance in session 1 was 3.61 meters, session 2 was 2.54 meters, and session 3 was 0.44 meters. There were significant differences between sessions 1 and 3 ($p = 0.03$) and sessions 2 and 3 ($p < 0.01$) using a T-Test for Independent Samples with Bonferroni's Correction.

These findings show that participants moved the most documents early on in the sensemaking process. Additionally, these early movements that the participants make are long movement distances. However, as the participant progresses further into their sensemaking, they make fewer document movements with smaller overall document movement distances. We believe this matches the overall sensemaking process of forming an evidence file, schemas, and re-representations of those schemas to formulate a hypothesis of findings. Overall, these results show the most refinement of the spatial layouts occurring during sessions 1 and 2.

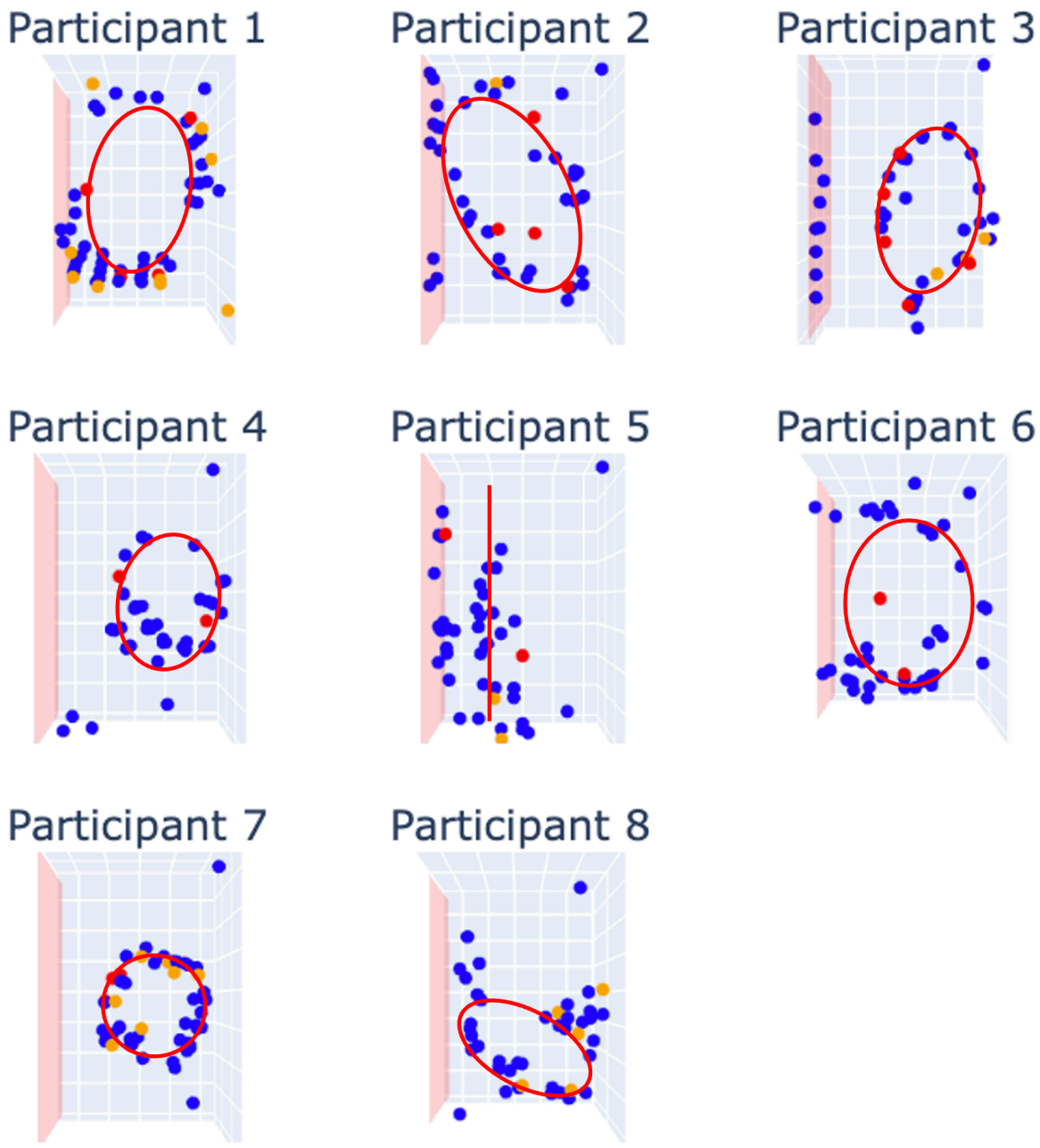


Figure 3.3: All participants' final spatial layouts formed during their analysis. These spatial layouts are portrayed using a top-down scatter plot. The red area on the left side denotes the bulletin board. Each document is represented using a blue dot, while notes are red and labels are yellow. The red curves approximate the overall organization structures formed by the participants.

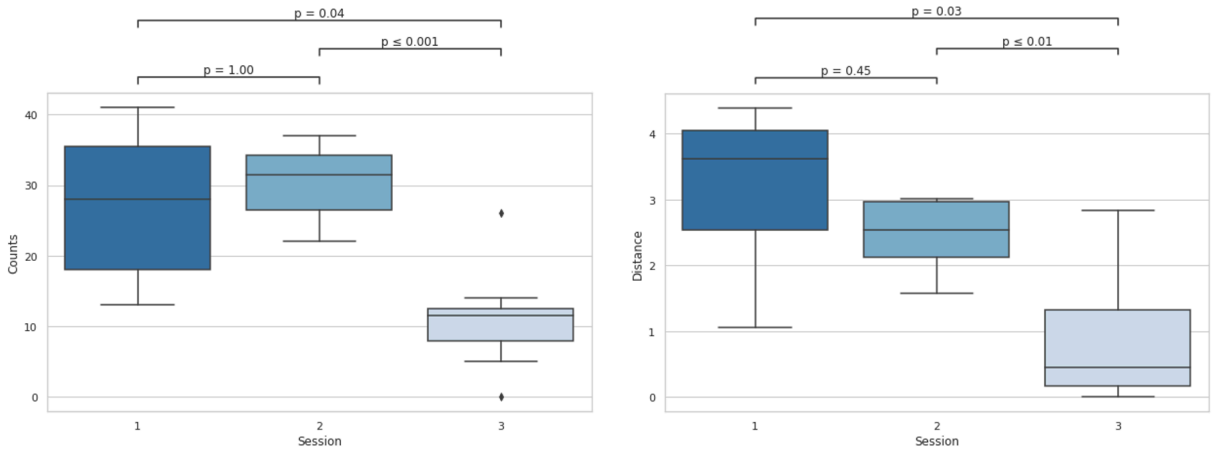


Figure 3.4: Number of documents moved during each session of analysis (Left). Distances (meters) that participants moved each document in each session (Right).

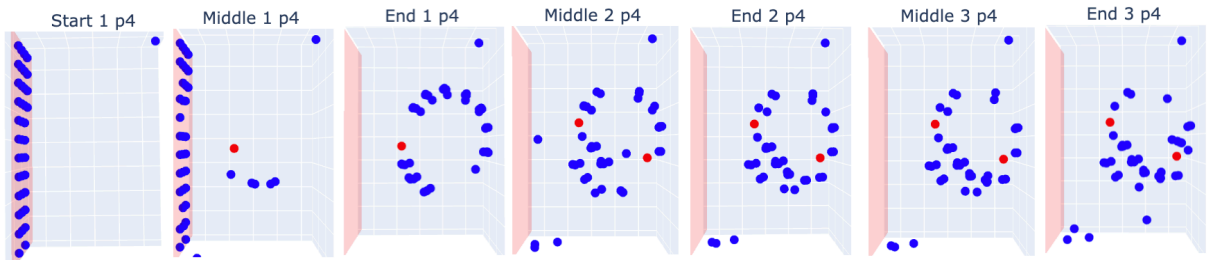


Figure 3.5: Evolving spatial organization for participant 4. From left to right we have the start, middle, and end (session 1) middle, end (session 2) and middle, end (session 3).

While we expected to see very few refinements occurring in session 3, when the participants were focused more on the report writing, we did find that they continued to make minor refinements. This demonstrates that even in the later stages of the sensemaking process, the schemas in the immersive space were still being refined.

We also analyzed the spatial organization at the beginning, middle, and end of each session for each participant. An example visualization (for P4) can be found in figure 3.5. Our primary finding from this analysis was that participants' spatial layouts became more refined over time. These refinements create more structure within the immersive space (e.g., turning a cluster into a timeline column (figure 3.7). This suggests that as incremental refinement occurred, the location of IST artifacts became more nuanced and purposeful. Using this change over time visualization, we are able to see these small refinements in the overall spatial layout made by the participants. In figure 3.5, for example, we are able to see a subset of documents that are pulled closer to the center of the cylinder from the start of session 2 to the middle of session 2. Additionally, we are able to see more small refinements

when we look from the end of session 2 to the end of session 3. While this is only one participant's change over time, we saw these incremental refinements of the spatial layout in all of the participants.

In addition to looking at the quantitative data collected, we also reviewed the interview questions: **1) How did you organize documents to extract meaningful information over the course of this analysis? And 2) Did your spatial layout change over the course of multiple sessions? If so, How?** The responses to the interview questions supported the notion of incremental refinement. For example, P4 stated:

> *"I guess on the first day, I organized all the documents into different organizations so CIA, FBI, then correspondence or sanctioned intercepts. On the second day, I started deriving all the different documents that I thought were useful, and then I just organized them into a timeline. I started connecting names from the different documents from the different organizations that I had previously organized on the first day. Then today [session 3], I just went off the outline I created yesterday and continued to connect names and relationships across all the different documents."*

Additionally, P1, stated that:

> *"At the beginning, it was all very scattered, and I was trying to make sense of all the documents and how they related to each other. By session two, I was starting to form clusters around the left side with relevant topics and then eventually settled on columns because they more compactly organized my thoughts and then created this layout. Session 3 was mostly the same because it seemed like this layout was working just fine, and I was able to extract the information quickly. Then all these other documents that I consider to be less relevant just stayed more or less scattered around here on the other side [right side of the space]."*

Lastly, we also looked at how, if at all, the overall organizational pattern changed over time. We found that 7 out of 8 participants who started with a given organizational pattern (i.e., semi-circular or environmental) ended with the same overall organizational pattern (with the exception of participant eight who started with an environmental layout and changed to a semi-circular layout as their sensemaking evolved). This could be due to factors such as the time limit of the task or inertia (participants only being willing to change the way they were analyzing the documents if there was a significant perceived benefit for doing so). We suggest additional longitudinal studies to understand how overall organization pattern changes from session to session as participants become more familiar with an immersive analytic tool. This would provide a deeper insight into the organizational patterns and sensemaking strategies users would use "in-the-wild" as they become more comfortable and even an expert with immersive analytic tools such as IST.

Other Spatial Organization Features

We were also interested in how the spatial layouts reflected the content and properties of the text documents, including relevance, plotlines, reporting agency, and chronology.

Relevant vs. Distractor Documents

One of the features of our text-based dataset is a set of distractor documents. To understand the spread of the documents, we calculated the average distance of all the distractor and relevant documents from the center of the space. Distractor documents were on average 2.60 meters from the center (SD = 0.798 m), while relevant documents were on average 2.24 meters from the center (SD = 0.844 m). We found that relevant documents were significantly closer to the center of the space ($p < 0.001$) using a T-Test for Independent Samples with Bonferroni's Correction.

We see a large standard deviation in our analysis, which could be due to the small sample size. However, this result still suggests that for some participants, important documents are organized closer to the center of the immersive analysis space, where the participant was placed at the start of their analysis. More analysis on relevance distance is still needed. If this result is confirmed, it might suggest that we could design semi-automated features to provide insight to users based on the documents closer to the center of the space.

Plotline Encodings

We also wanted to see where documents that belong to a specific plotline appear in relation to other documents in that plotline. In other words, we wanted to evaluate the clustering that may occur between documents related to the same topic. Details on the dataset's plotlines can be found in section 3.3.4.

Using the plotline encodings, we generated plots similar to those seen in figure 3.5, which illustrated how each plotline's spatial organization changed over time. We found that while documents within plotlines tend to be clustered together, not all documents from a single plotline end up in a single large cluster. Often the documents within a plotline are organized into multiple small clusters organized around the immersive space. Additionally, we found that the plotline clusters are not clearly defined. Documents from different plotlines are often mixed in with each other in the clusters formed by the participants. Lastly, as participants get further into their sensemaking process, the small clusters of documents related to the same plotline become more defined, with fewer distractor documents mixed in.

Reporting Agency Organization

Another organizational strategy that we looked into was if the reporting agency (CIA/FBI/SI) affected the spatial structure formed by the participants. We found that for one of

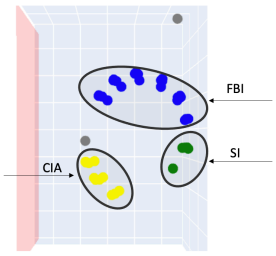


Figure 3.6: Reporting agency of each of the documents in P4's spatial layout after session 1. Yellow represents CIA, Blue represents FBI, Green represents SI, and gray is the prompt/note.

the participants, P4, the reporting agency largely affected their spatial layout. After day 1, they had sorted each of the documents by reporting agency into a sphere around them, as reported in their interview (section 3.4.1) and as seen in figure 3.6.

After this organizing by the reporting agency, the participant then began to move documents in the same plotline closer to each other and out of their initial organizational clustering. While we only had one participant organize their documents by reporting organization, this strategy of developing their organizational structures could be used to inform future design ideas for IST and similar immersive analytic systems. One feature that could be developed is an automatic sorting feature in which users can sort based on different features of the dataset.

Timelines

Another common strategy we found that the participants used for organizing their documents was to use chronological order. Often, these timelines were formed in columns within the space. In total, six of the participants organized things in either chronological or roughly chronological order within their final spatial layout. An example of a column-based timeline formed by the participants can be seen in figure 3.7. Additionally, P4 and P1 referenced their column timelines in the interview responses in section 3.4.1.

Since this task was focused on reports from specific dates, we believe that the participants were forming these timelines to develop an understanding of the evolution of the events. This suggests we could add additional features to our system to help support the user during the sensemaking task. A built-in timeline generator might be useful for participants in a task where time is a relevant factor, such as intelligence analysis or news analysis.

Trash Pile

Finally, participants utilized trash piles (clusters of irrelevant or unwanted documents) in their spatial organization. These piles are often located in the extremities of the immersive space. Some of the common places for trash were in the corners or along the sides of the tracked area. An example of a trash pile can be found in figure 3.5 in the bottom left-

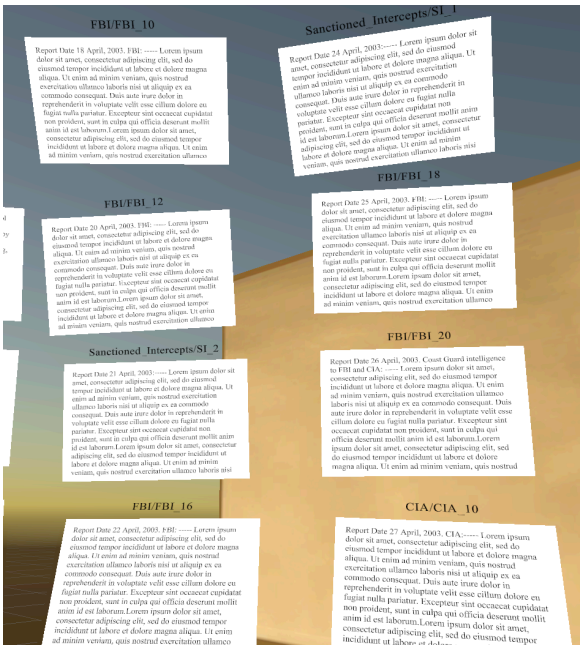


Figure 3.7: Two columns that are organized into timeline order, on the left the top document is posted on April 18, and bottom on April 22. The right-hand column goes from April 24 (top) to April 27 (bottom)

hand corner of the plots starting in the middle of session two. Additionally, trash piles are referenced by P1 in their interview response seen in section 3.4.1. These trash piles were found in five of the participant's spatial organizations. In addition to creating trash piles, three participants placed documents that were not needed in their analysis back on the bulletin board. The use of bulletin board and trash piles were not mutually exclusive.

Overall, in our analysis of RQ1.1A, we found that participants formed similar spatial patterns to those seen in [60]. We also found that the spatial layouts change over time, with most refinement happening between sessions 1 and 2 and smaller refinements happening between sessions 2 and 3. Additionally, we found many different sub-organizational strategies within the spatial layouts; these included plotline organization, relevant/distractor document organizational patterns, reporting agency organization, timelines, and trash piles. We do not claim this is a comprehensive list of all strategies used by the participants. However, it is clear that variety in organizational strategies is possible because IST provides great flexibility and freedom of choice for completing a dynamic and complex task such as intelligence analysis.

3.4.2 RQ1.1B: Mapping of Spatial Layout to Report Structure

To address question RQ1.1B, we were interested in how the organizational layout formed within IST mapped to the reports that were generated by the participants.

One approach to understanding the mapping was through the interview question: **Does your spatial layout reflect the organization of your report? If so, how?** Two participants mentioned a clear organization that was reflected in their report, such as P1 stating, *"Yes, I started right here, and then, I just kept moving this way [Left → Right].... with a very natural flow of information."* Additionally, two participants reported a rough mapping between spatial layout and report, such as P5 stating, *"It would mainly show the documents that I looked at most, it is a bit messy, but I think the information I used in the report will be closest to me or the desk"*. Four participants said they did not use any clear mapping between the two, such as P8 stating, *"Not actually, I think maybe each label would be one of the findings"*.

Next, we looked at the citation order of the documents within the written reports and compared the order to the location of the documents in the immersive space.

For each participant, we developed a document reference order, which was, on average, 13.75 documents long (SD = 2.45). Using this reference order, we drew a line through their immersive spatial layout to understand the transformation of the 3D space to the 1D hierarchical reports. Additionally, we were interested to see patterns between documents cited in a paragraph together. An example of the scatter plot with the reference line and paragraph coloring can be found in figure 3.8.

We calculated the length of each reference path starting at the first document cited in the report and tracing the line to the last document cited in the report; we also normalized this length by the total distances between all documents in the immersive space (accounting for spread of documents). Where we were expecting to see shorter path lengths correlated to quality of the reports (section 3.4.5). Overall, we found participants had an average length of 21.77 meters (SD = 9.58) on the reference path with a normalized value of 0.13 meters (SD = 0.04). While we were expecting to see clear transformation paths from the spatial layouts to the reports, we only found patterns for four participants. These patterns (either roughly Right → Left or roughly Left → Right where orientation is represented by the facing the bulletin board) had few intersections on the reference line. An example of the roughly Left → Right pattern can be seen in figure 3.8. The remaining four participants had no distinct pattern to their transformation path. This could be due to participants having freedom in organizing their immersive space and potentially not thinking about a linear report while creating the spatial organization. Such layouts would not always correlate with a clear report structure. Additionally, the tool was unfamiliar to users, and they may need to practice to fully develop organization patterns that make the reporting stage of the sensemaking process flow seamlessly. Future studies with more expert participants could be conducted to further evaluate the mapping between spatial layout and report.

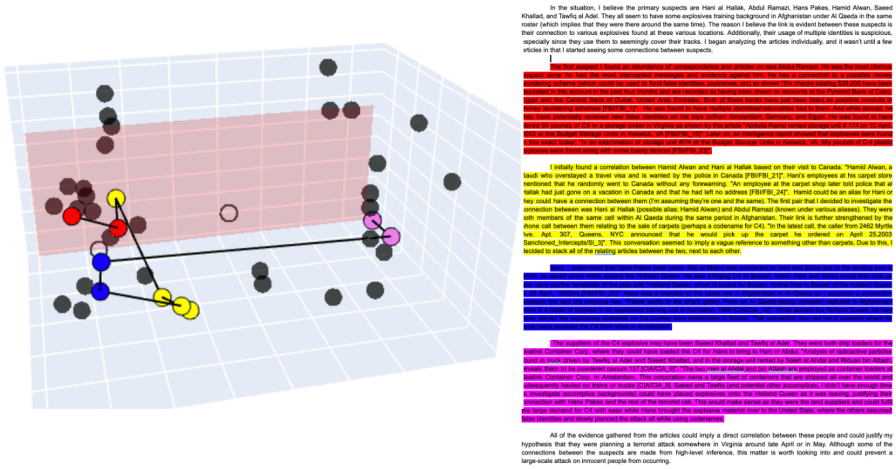


Figure 3.8: 3D scatterplot of P6's spatial layout with the document reference order traced. Gray documents were not referenced in the report. The red rectangle represents the bulletin board, and the open gray circle represents a free-standing note. Colored documents were referenced in the report coordinated with the paragraph coloring on the right.

We were also interested to find whether documents that were cited near each other in the report tended to be close together in the spatial layout. We would expect to see this due to clustering based on similar information; however, we were surprised to see that documents that are cited within the same paragraph do not always appear next to each other in the immersive space. As users become more familiar with the tool, we might see more organization based on report structure. Overall, the relationship between the immersive spatial layout and the report is more complex and needs further study.

3.4.3 RQ1.2: Physical Navigation

To address RQ1.2, we examined how participants navigated the virtual space while conducting both their analysis and their report writing.

Participants' movement through the space was reflected in the log files. We smoothed the data from the logs to get one log per second. Using this data, we calculated the overall distance that each participant moved during each session. We then normalized the distance walked by the total time spent in the headset. Overall, we saw no difference between the normalized walking distance between sessions 1 and 2, or sessions 1 and 3, but we did see a statistically significant difference between sessions 2 and 3 ($p = 0.02$) using a T-Test for Independent Samples with Bonferroni's Correction. See figure 3.9.

We also looked at both the trajectory and the dwell time (time spent in each region of the workspace) as seen in figure 3.10. The overall trajectories revealed two categories of

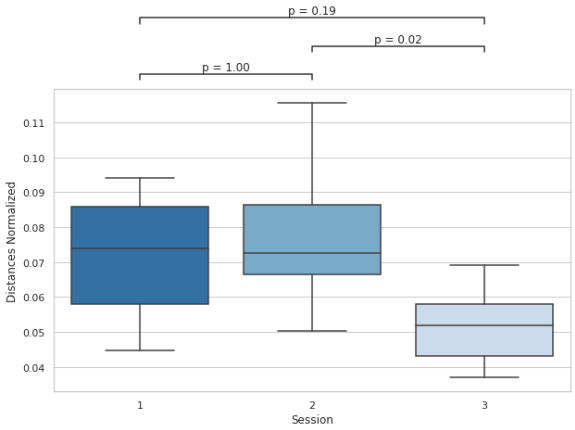


Figure 3.9: Participant walking movement normalized by time spent in headset per session.

movement patterns. Four participants navigated through the majority of the space provided during their analysis, while the other four participants remained mostly stationary during their analysis. In the figure, we can see that the participant (P1) navigated the majority of the space provided, and as the sensemaking process evolved (session 2 → session 3), we see less navigation of the space and more time spent in one location (figure 3.10), as indicated by one dark green area. We saw this pattern in the four participants who navigated the space during their analysis. The remainder of the participants had similar dwell times across all three sessions, which was largely due to minimal walking movement around the overall space.

While we did see an overall downward trend in the number of places dwelled from session 1 to 3, the participants still utilized the space to reference information far from the dwelled location, as suggested by the trajectory lines, which span the majority of the space inside the spatial layout of documents. This suggests that the ability to reference information organized in the space is not limited to the early stages of the sensemaking process.

Similarly, we analyzed participants' use of the wheeled table and keyboard during their analysis. Overall, we found that the four participants who navigated the space during their analysis also moved the keyboard around. The participants who tended to stay stationary also had a primarily stationary keyboard. These results matched the findings from the participant's navigation across sessions.

Overall, we found that participants navigated the most space during sessions 1 and 2, while in session 3, the participants were more stationary due to writing the report. However, while more stationary, the participants still navigated through the space to use the reference material around them. Our findings suggest that participants continue to search for support and reevaluate their hypotheses while completing the presentation stage in the sensemaking process.

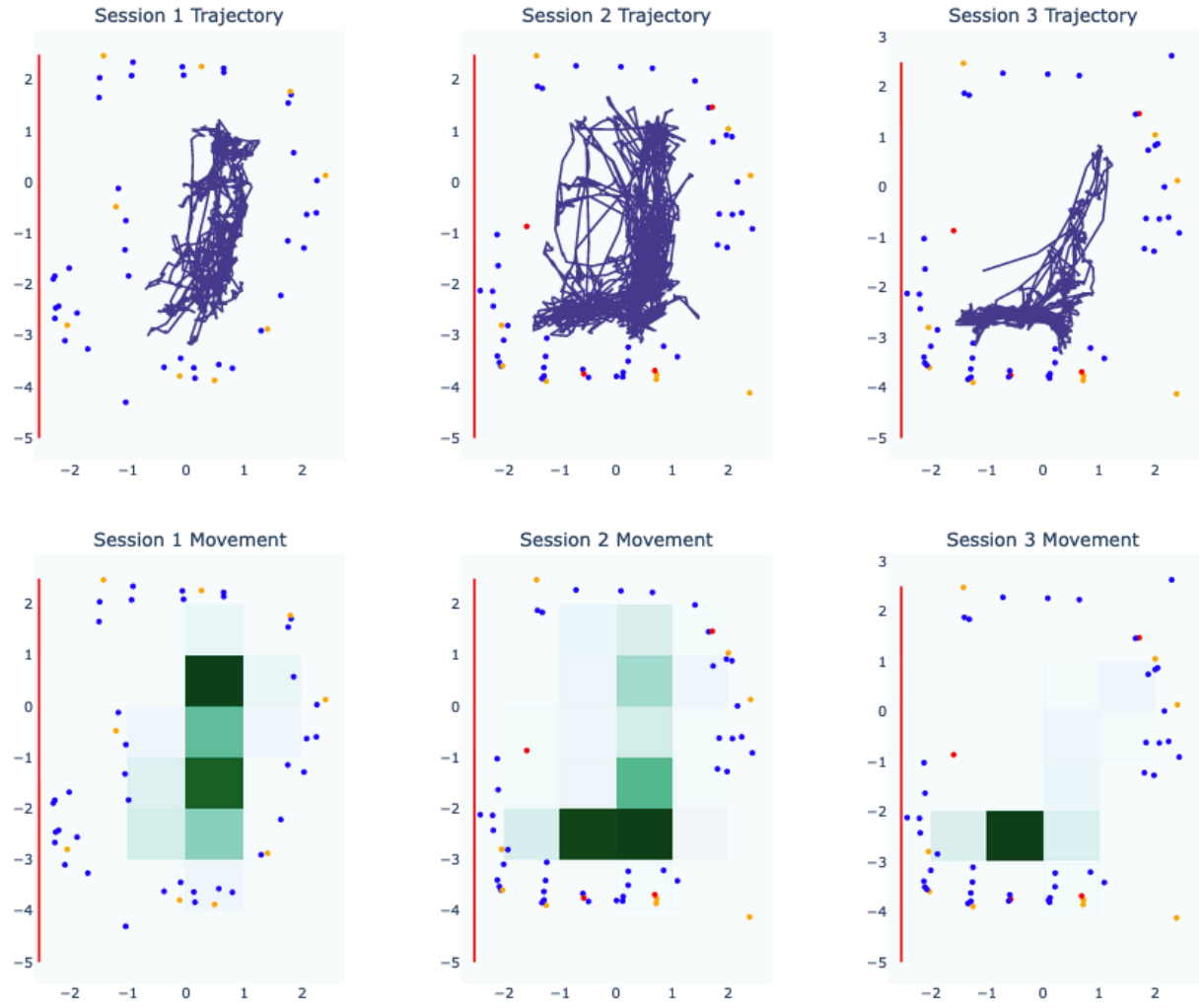


Figure 3.10: Trajectory of movement for P1 during each session (top row) and the dwell times of the participant across each session (bottom row) where a darker color indicates more time spent in that location. Each plot has the document locations at the end of the session represented with blue (document), yellow (label), and red (note). The red line denotes the bulletin board.

3.4.4 RQ1.3: Bridging Technologies

RQ1.3 asked how participants would make use of immersive and traditional technologies simultaneously to complete the task. We observed three strategies for bridging the technologies during the report writing stage:

1. **Reference**: Entering and exiting the immersive space multiple times by lowering or lifting the headset in order to reference information in the virtual environment as needed during report writing. In the reference technique, the participants would switch between the immersive space and Google doc rapidly to access information at the time it was needed.

2. **Export**: Creating a note in the immersive space with the information that the participant needed for the report and exporting that note into the Google doc using the copy and paste feature. The export was only conducted once after all details needed for report writing were offloaded into the IST note.

3. **Hybrid**: A combination of the two techniques where information was offloaded to or taken from an IST artifact and moved into the Google doc using the copy and paste feature on multiple occasions during the report writing stage.

Six participants used the reference technique, one used the export technique, and one participant used the hybrid technique for the report-writing process. While raising and lowering the VR headset repeatedly is not ideal, this probably reflects the perception that creating content directly in the Google Doc using familiar tools was preferable to creating content in the immersive environment and then exporting it. While there are current limitations in the technologies available, using augmented reality (AR) or a pass-through view of the laptop within the virtual environment could provide the benefit of spatial organization in the world around the user as well as the benefits of easy bridging between traditional and immersive technologies, especially for users who utilize the reference strategy.

3.4.5 RQ1.4: Report Scores and Strategies

Quality To grade the reports for quality, we worked with a set of four professional intelligence analysts. Before grading the reports, the experimenter met with the analysts to onboard them to the grading process. They were then each provided the rubric in a Google Forms format, a document that specified the order in which to grade the reports (randomized for each grader), and a PDF copy of each of the participants' reports. The analysts were given a total of two weeks to complete the grading.

Correctness Using the correctness rubric, the experimenter graded each report generated by the participants. Reports were graded in a randomized order.

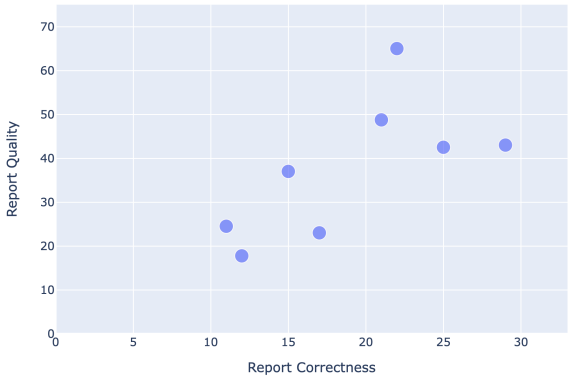


Figure 3.11: Report Correctness score vs. Report Quality score.

Scores

There was a total of 75 points possible for quality. We had an average score of 37.69 (SD = 15.62). For correctness, there were 33 points possible, and we had an average score of 19.13 (SD = 6.17). Figure 3.11 plots the quality and correctness scores for each participant. There is an overall positive correlation between correctness and quality scores using Pearson's correlation coefficient; however, these results are not statistically significant ($p = 0.07$).

To understand what factors might influence report correctness/quality, we ran correlation analyses between the report scores and various interactions performed by the participants (although we understand the limited statistical power of this approach due to a small sample size). The interactions we looked at were the number of highlights, notes, searches, distance traveled, labels made, highlight length, note length, time spent grabbing documents, and number of documents grabbed per session. We saw a significant negative correlation (-0.77) between the number of searches made during session three and the overall report score using Pearson's correlation coefficient ($p=0.015$). This means that using more searches in the report writing stage was linked with a lower overall score, suggesting that participants who relied on the search tool instead of externalized memory in the last session performed worse on the reporting task. None of the other correlations were statistically significant.

We also investigated possible correlations between report score and other features, including spatial organizational pattern (environmental/semi-circular), the shape of the reference path, length of the reference path, the overall spread of spatial layout (total distance between all documents in immersive space), distances between documents within a plotline, number of distractor documents referenced in the report, and number of topic switches in the report (switches between plotlines/distractors). We found a significant positive correlation (0.72) between the reference path shape and the overall report grade using Pearson's correlation coefficient ($p = 0.043$). This suggests that the more organized the reference path of the spatial layout to the report, the better a participant scored on the report rubrics. None of the other correlations were statistically significant.

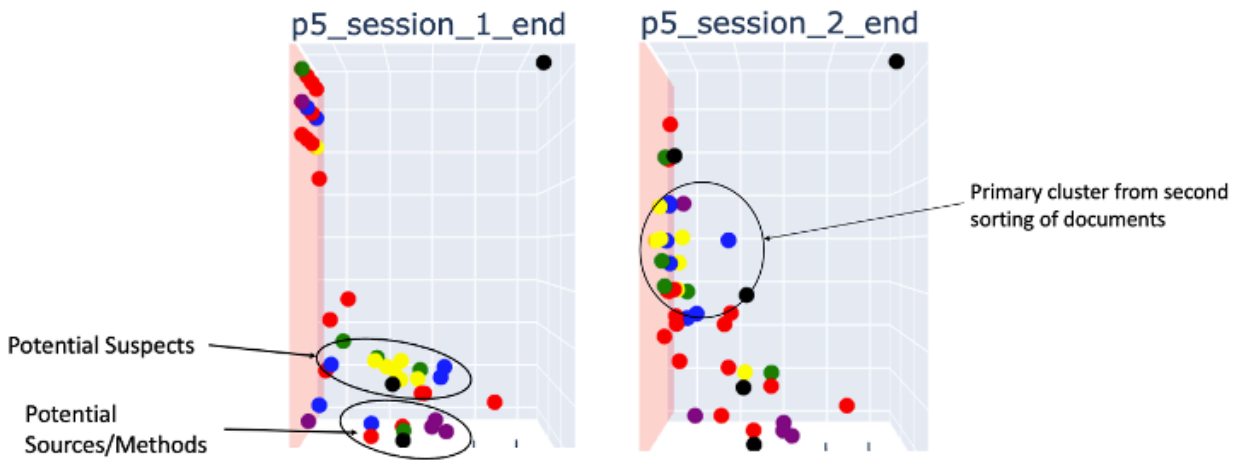


Figure 3.12: P5's Spatial Layouts: end of session one (left), and end of session two (right). The colors represent plotline encodings where red represents distractors, blue/green/purple represents plotlines, yellow represents coordination, and black represents notes/labels/prompt.

Finding few significant correlations between the interactions/features of the participants and the overall grades received on the reports could be due to the small participant pool (resulting in limited power for statistical analysis) or the diversity of strategies used by participants. The IST system may provide users with the flexibility to complete the task well with various analysis strategies. However, we recognize that by running many correlations in our data, we were more likely to find something that was positively or negatively correlated. We report these findings as an interesting direction for future analysis of this type of data, though future work will be needed to verify and evaluate correlations.

Overall Participant Strategies

Using screen-captured videos of analysis sessions, we developed a narrative description of the overall strategy that each of the participants used for their analysis. In this section, we highlight the strategies of our two highest and two lowest report-scoring participants based on both quality and correctness.

Highest Scorers' Strategies

Participant 4 began his analysis by reading the content of the first few documents on the bulletin board. After analyzing the content of the documents, he organized the documents into 3 different categories in a sphere around him. These categories aligned with the three

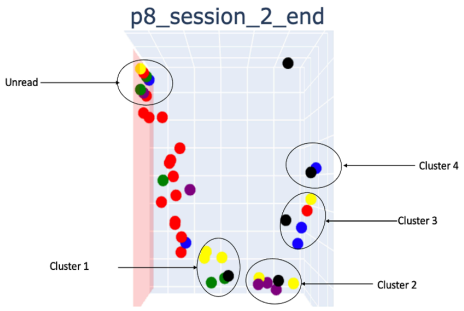


Figure 3.13: P8's spatial layout after the end of session two. Color encoding is the same as figure 3.12.

different reporting agencies as seen in figure 3.6. Additionally, the participant placed the documents into chronological order within their categories. Once the categories were established, the participant created a note capturing relevant names and aliases for all the documents within a category. During the development of the notes, the participant moved documents with key characters in between two clusters (creating a new cluster). Additionally, the participant used the search to find more documents relating to the key characters, which were moved into the new cluster. During the report-writing stage, the participant used the reference strategy to incrementally get information from the immersive space into the written report.

Participant 5 began her analysis by sorting the documents provided. She would select a document off the wall, read it, and place it into the environment around her. Once she gained enough information on the dataset, she created labels to help organize the sorted documents as seen in the left side of figure 3.12. As she continued with sorting, she cross-referenced information within the groupings based on people, places, and phone numbers. After completing her initial sorting, she started cross-referencing documents between the groupings and moved documents into a new cluster in the center of the room (second sorting) as seen in the right side of figure 3.12. Once the documents were sorted a second time, small refinements of document location occurred, and highlights were applied to some more relevant documents. During the report-writing stage, she utilized the reference strategy to incrementally get information from the immersive space into the written report.

Lowest Scorers' Strategies

Participant 7 started his analysis by taking documents off the wall, reading them, highlighting names, places, and phone numbers, and placing them in the immersive space. When the number of names became cumbersome, he created a note in which he kept track of all names seen within the documents. A small amount of cross-referencing between documents occurred before placing the document into the space. Once all of the documents were off the bulletin board and sorted around him, the participant used the search feature to find connections between documents. During the final session, he continued with analysis and

spent time focusing on offloading information into an IST note. This participant used the export strategy to bridge between the immersive and the traditional space.

Participant 8 began her analysis by selecting documents off the wall, highlighting names, places, and phone numbers, and putting the documents back onto the wall. Additionally, during the initial reading, she would offload some information into the notes attached to the documents. After reading the majority of the documents, she created some distinct clusters on one side of the room. By the end of session two, she had created four distinct clusters, and labels were applied to each of the clusters as seen on the right side of figure 3.13. There were still some documents that were left unread on the bulletin board at the end of session two. During the final session, the participant continued to refine her spatial layout, and during the report-writing stage, she used the reference strategy to access information within the immersive space.

One commonality between our two highest scorers was creating meaningful organizations early on in the sensemaking process and refining those structures as knowledge was learned. Another difference between our two highest scorers and the lower scorers was when highlights were made. Our lowest scorers tended to highlight a lot of information during the first reading of the documents. We believe that this may have led to distracting information being highlighted as important and, therefore, distracting the participant from the main storylines.

We have chosen to share the overall strategies for these participants not just to highlight some commonalities/differences but to reinforce how immersive analytics systems provide flexibility in completing sensemaking tasks. In our study, we had eight different participants, which led to eight different strategies for analysis! While there are similarities within participants, each user was given the flexibility to complete the analysis in a way that worked for them.

3.5 Limitations

Due to the COVID pandemic, we had a relatively small participant pool, which limited our power in statistical analyses. In addition to this, COVID restrictions caused three participants to have two sessions in a single day separated by a minimum of three hours. This could have affected the distributed sensemaking analysis design.

Another potential limitation of our work is that intelligence analysis is a complex task, and the participants were not trained specialists in this area. This may have led to low analysis scores and different organizational patterns than we may expect to see with professionals. More work would need to be done to understand the strategies that professionals use in an immersive system such as IST. Another limitation of this work is the text-entry system. While we provided the users with the ability to enter text using the wheeled keyboard and table, the participants still struggled to see their fingers on the keyboard despite our efforts

to improve visibility by removing some of the face cushion on the lower portion of the VR headset.

We also acknowledge the potential effects the study design may have had on our findings. In order to have the participants produce both an outline and a report of their findings for our analysis, we had to provide checkpoints for our users to achieve in each session. This could have affected their overall analysis strategy. To combat this limitation, we provided our users with guidance that they were free to begin or return to stages or checkpoints as needed over the course of their analysis.

3.6 Conclusions

This study examined how the IST approach supports the sensemaking task from foraging to reporting. Our analysis found that participants formed both environmental and semi-circular spatial organizations, and these organizations are incrementally refined in the later stages of the sensemaking task. The refinements that occurred late in the sensemaking process include a few documents moving a short distance compared to the movement patterns seen in sessions one and two. Additionally, we saw the most navigation of the space occurring in sessions one and two and less movement with more dwell times in session three. Next, we found that the participants used three different strategies to bridge the immersive space and the traditional technologies (laptop and text editor) during the report writing stage. These strategies included reference, export, and hybrid.

In the analysis of the reports developed using IST, we compared the spatial organizations and the report structures. We found that half of the participants had a roughly linear transformation path with few intersections along the path. We found an overall positive correlation when comparing these transformation paths to the overall report grade. Additionally, we found a negative correlation between the number of searches within the third session and the overall report score when evaluating interactions and report scores. Lastly, when assessing the overall strategies that the participants used, we found that the IST approach provides flexibility in the method used for analysis.

We believe the results from our study show promise for immersive analytic systems, such as IST, in providing flexibility in text-based analytic tasks. However, one limitation of this study was that it only evaluated how novice users sensemaking within IST. By understanding the strategies and organization professional analysts use, we could develop immersive analytic systems that provide better support during the sensemaking task.

Chapter 4

Investigating Professional Intelligence Analyst Sensemaking in Immersive Space to Think

In the last chapter, we presented an exploratory study on how novice users utilized IST for a multi-session sensemaking task. This study gave us a deeper understanding of how these systems can be used for the sensemaking process. However, extending these findings to professionals engaged in daily sensemaking remains uncertain. To bridge this gap, we conducted a study involving professional intelligence analysts to comprehend how these systems could be used during a sensemaking task within a professional user group. This study also addresses Research Question 1 (RQ1): How do sensemaking strategies evolve during multi-session analysis in Immersive Space to Think?

4.1 Introduction

In the world of big data, there is a growing need to provide better ways of filtering and sorting the large amounts of data presented to us. In addition, providing analysis spaces for understanding the presented data will be critical to parse through the data to find patterns and develop key insights. With the advent of new lower-cost immersive technologies, there are new opportunities for using these technologies to build an analysis space of the future for completing sensemaking tasks. Immersive Analytics (IA) is an emerging research field combining previous research on visual analytic systems and data visualization with new immersive technologies such as virtual/augmented reality interfaces. These immersive technologies provide richer sensory information, including better depth cues, increased spatial orientation understanding, better peripheral awareness, and increased engagement compared to traditional desktop displays [1, 10, 68]. Additionally, they build on the ideas of distributed cognition [44], offering an immersive way to externalize cognition into the environment around you.

With this growing research area, there has been a large body of work dedicated to understanding how users work within these spaces with both quantitative data using 3D visualizations [6, 17, 96] and non-quantitative data such as text-based analysis [59, 60]. Many

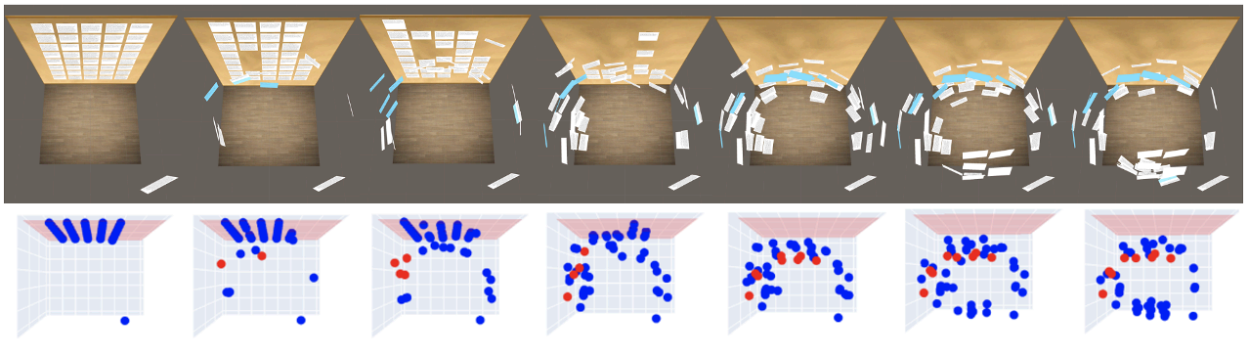


Figure 4.1: High-Level View of an Immersive Space to Think Analysis. This example is of Participant Five's analysis over time. From left to right: start, middle of session 1, end of session 1, middle of session 2, end of session 2, middle of session 3, end of session 3. Top Row: Top-down orientation into the immersive space. Bottom Row: top-down view of 3D scatter plots used for data analysis in this work. The blue dots represent the documents provided to the users for analysis, and the red dots represent notes/labels created by the user during analysis.

of these non-quantitative data-based studies have focused on understanding how these immersive analytic systems can support the sensemaking process [73] including single-session analysis [59, 60], multi-session analysis [19] and trying to understand how strategies used and structures made within these systems in these systems change as users become more familiar with their usage [20]. These studies have provided many insights into how IA systems can benefit the sensemaking task.

However, there is a large gap in our understanding of professional users within IA systems. Some research has been dedicated to using professional users within immersive analytic prototypes to understand how professional economists explored economic data [6]. However, to our knowledge, a study has not yet been done on how professional intelligence analysts utilize an IA system during an analysis task. Additionally, we are unaware of work that aims to compare professional and novice analysts within these systems. This paper seeks to fill that gap by providing a deep data analysis on professional intelligence analysts' usage in an IA system - Immersive Space to Think, as well as providing insights into how these professional intelligence analysts differed or were similar to novice user populations studies before.

The key contributions of this work are as follows:

- A deep understanding of professional intelligence analysts' multi-session sensemaking process in immersive space
- A detailed comparison of novice vs. professional analysts in immersive sensemaking
- A validation of previous findings presented in IA sensemaking systems

4.2 User Study

4.2.1 Goals and Research Questions

A professional intelligence analyst uses analytic thinking skills to work on complex sensemaking tasks daily. These professionals are trained to complete these complex tasks and develop a unique set of critical thinking skills to achieve their goals. To fully understand how the IST concept can support the sensemaking task, we believe that evaluating professional intelligence analyst sensemaking within this system can provide us with new insights on how these tools would be used "in-the-wild". This user study aimed to understand how professionally trained intelligence analysts complete sensemaking within an IA system and gather feedback from those analysts on future design ideas for IA prototypes. This study also aims to understand how professionally trained analysts and novice users differ in IA system usage, as the type of strategies and organizational tools used by these analysts may differ from those identified in previous work.

To compare these two populations, we rely on the data and user study design as seen in chapter 3.

Our research questions are as follows:

Spatial Structures

- ***RQ 1.5A*** What spatial structures do professional analysts form in IST?
- ***RQ 1.5B*** How do these spatial structures change over multiple sessions?
- ***RQ 1.5C*** How do these structures compare to novice users?

Transformation Path

- ***RQ 1.6A*** How do the spatial structures formed map to the 1-dimensional written reports?
- ***RQ 1.6B*** How do these transformation paths compare to novice user transformation paths?

Physical Navigation

- ***RQ 1.7A*** How do professional analysts navigate the space while completing the sensemaking task?
- ***RQ 1.7B*** How do professional analysts compare to novice users in physical navigation while completing a sensemaking task?

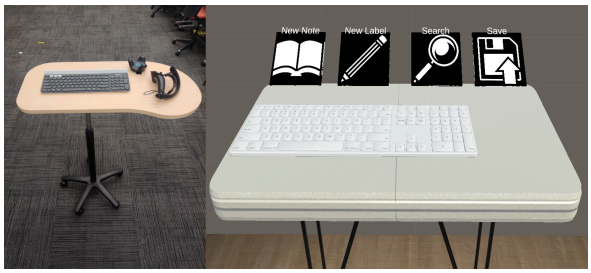


Figure 4.2: Left: Text Entry System. A HTC Vive Tracker on the wheeled table, a wireless keyboard, and space for the controller to be set down while typing. The keyboard was replaced with a 2019 Mac Book Pro using Google Docs during report writing. Right: VR Headset view of the tracked keyboard and table. There are four menu buttons attached to the keyboard. These buttons are New Note, New Label, Search, and Save.

Strategies

- ***RQ 1.8A*** What strategies do the professional analysts use, and how does this correlate to the quality of the written reports?
- ***RQ 1.8B*** How do these strategies change over the analysis sessions?
- ***RQ 1.8C*** How do these strategies compare to those of novice users?

4.2.2 Apparatus

A desktop computer was used to run the IST system using an HTC VIVE Pro [1] head-worn display with a wireless adapter to allow for a tether free walking in the space. The participants held one Valve Index wireless controller in their dominant hand for interacting with the system. The 3D interaction technique used in this IST prototype was Ray-Casting with reeling. The document size is 0.5x0.3 (m), which supports about one paragraph of text (Text-Mesh Pro text size at 40) without the need for scrolling. Participants used a wheeled keyboard table with a wireless keyboard for text input. Small amounts of foam were cut out of the VIVE Pro headset to provide better support for text entry so the participants could see their fingers on the physical keyboard. A 2019 MacBook Pro with Chrome Remote Desktop connected to the desktop PC was used for the report writing. An example of the wheeled table used for text entry and report writing can be seen in figure 4.2. A Steam VR 2.0 Lighthouse tracking system was used for a tracked area of 3x3 meters. In the tutorial phase of the study, the participants were trained on the size of the tracked space and were also shown that the virtual floor represented the boundary of the tracked area. In addition, the experimenter watched to ensure that the participant did not get too close to the physical

[1] https://www.vive.com/eu/product/vive-pro/

room boundaries. This study's amount of tracked space differed from the prior chapter, where the participants had 4 x 8 meters. This was due to the constraint of recruiting professional analysts at a location convenient to them. We highlight this as a limitation and apply normalization when necessary in our analysis to draw comparisons between the two groups.

As this study aimed to evaluate the differences between Novice users and Professionally Trained analysts, the features used in this IST study were the same as in chapter 3.

4.2.3 Experimental Task

This experiment aimed to understand how professional analysts complete sensemaking in an immersive analytic prototype. Additionally, we aimed to evaluate how professional and novice users differ on structures and strategies used to inform future design ideas for immersive analytics systems. For this comparison, we utilized the same experimental task as seen in [19]. This task asked the participants to analyze a set of text-based documents, of which our participants would have no prior knowledge, to write a report of their findings at the end.
The task provided to the participants was as follows:

Today is April 27th, 2003. You are an intelligence analyst working for the federal government. It is believed that terrorists are planning an imminent attack on the United States. Other analysts have gathered a set of potentially relevant documents containing information about potential suspects. These documents have been loaded into the Immersive Space to Think system for your analysis. Your goal is to analyze the information and develop a specific hypothesis about any potential planned terrorist attack(s) against the US. Your hypothesis should identify ***who, what, when, and where****.*

Over the course of 3 sessions, you will develop a report for the Office of Homeland Security. To do this effectively, you must prepare a defensible and persuasive report that describes exactly what your conclusions are based on the documents provided. During session one, you will focus on analyzing the documents and preparing a specific hypothesis of your findings (who, what, when, and where). During session two, you will continue your analysis and prepare a written outline of your findings. During session three, you will finalize your analysis and write a 1-2 page report of your findings. Your report must state what action or actions the terrorist(s) are planning, where they will occur, and when they will occur, using evidence and citations from the documents provided. Our hope is that your report can be used to thwart the terrorist(s) threat.

Participants were notified that the milestones for both the outline and report were designated as goals to be fulfilled by the conclusion of sessions two and three. They were also informed that they could progress ahead or revisit the checkpoints as necessary to support their

sensemaking process. Our analysis identified participants who proactively completed the checkpoints ahead of schedule and returned to checkpoints as needed to support their process.

4.2.4 Dataset

Again, since this study aimed to understand similarities/differences of sensemaking between professionals and novices in our prototype IST, the dataset used in this task is the same as seen in chapter 3. This allowed us to compare both the strategies used for completing their sensemaking task and the overall spatial structures that were formed as the task was a controlled variable between these two separately studied.

This dataset, Sign of the Crescent, is a fictional intelligence analysis dataset containing 40 text-based documents. Each document is about a paragraph in length. The dataset contains many features, including names/aliases, addresses, phone numbers, bank account information, places of business, passport numbers, etc. Additionally, this dataset contains relevant (23) and distractor (17) documents that add complexity to the analysis task.

The dataset contains three major plotlines related to a set of coordination documents. Plotlines boil down to the locations where the events related to the plotline are happening (Atlanta, Boston, or New York). Plotline 1 contains five documents, Plotline 2 contains seven documents, Plotline 3 contains four documents, and there are seven coordination documents. Along with these features, the documents purported to come from three different reporting organizations: FBI, CIA, and Sanctioned Intercepts. There were 24 FBI documents, 11 CIA documents, and 5 Sanctioned Intercept documents. During the analysis, all of the documents were provided to the users at the start of the session for analysis.

4.2.5 Participants

Due to the challenges of scheduling professional intelligence analysts for a user study, we had a relatively small pool with N equal to 11. The analysts were US Department of Defense employees and had an average of 9.72 years of professional intelligence analysis experience, with a standard deviation of 5.12 years. To protect the identities of the participants, age was not collected. However, we view years of analysis experience as a good proxy for age, with the age range representative of the US working-class population. We had six male and five female participants with the following VR/AR experience levels: Two participants had no previous experience, six participants had tried VR/AR once or twice, two participants had tried VR/AR 3-10 times, and one participant had tried VR/AR more than ten times. All participants had normal or corrected vision (glasses or contacts). The participants were recruited using email and word-of-mouth recruiting. The Virginia Tech institutional review board approved this study to ensure participant protection.

4.2.6 Measures

Before the study began, participants were guided in the consent process and informed that they could withdraw their consent at any time, in which case, any data collected from them would be deleted. After completing the consent process, the participants were asked to complete a **pre-study questionnaire**, which collected background data such as gender, analysis experience, VR/AR experience, general sensemaking strategies, etc.

Log Files - Collected during each session, and containing the participant's head movement, controller movement, keyboard movement, as well as any system-level interactions with the system (i.e., new notes, new label, grabbing document, hovering over a document, highlighting, etc.). All movement data was reported at about ten times/second.

Save Files - Generated once a minute during the session and contains the location of each IST artifact (documents, notes, and labels). We can create minute-by-minute snapshots of the immersive space over the entire sensemaking task using the save files.

Screen Recordings - Taken from the participants' first-person point of view during their session.

Interviews - After each session, interviews were conducted on session-specific questions, and after the study, the participant took part in a post-task interview.

Outlines/Reports - Generated by the participants during the experimental task. The Outlines were generated during session two, and reports were generated during session three within a Google Doc using the laptop placed on the tracked table.

In addition to the data collected for this study, we graded the reports for correctness and quality. For correctness, we evaluated the reports based on the ground trust of the dataset using a rubric created to look at the *Who, What, When, and Where* questions. The rubrics were developed with non-participant professional intelligence analysts. The rubric was designed to produce subjective quality scores, which measured conciseness, persuasiveness, clarity, completeness, relevance, and bias. These rubrics are the same ones used in chapter 3, and a copy of the report grading rubrics can be found in the appendix.

4.2.7 Procedure

This study follows the same procedure as seen in chapter 3. The study was broken down into three sessions, which are detailed below.

Session 1

Pre-session - Before participant arrival, all equipment used during the study was sanitized to help prevent the spread of virus. Upon arrival, the participants were asked to review the

consent document and ask if they had any questions. Upon receiving the consent of the participants, the study began with the background questionnaire.

Tutorial - Before entering the headset, the participants were trained on the Valve index controller and HTC Vive Pro adjustment mechanisms. After this, the participant would don the headset for the tutorial on how to use the IST system. The tutorial began walking around the immersive space to understand the physical boundaries of the tracked area. Then, using a tutorial dataset, the participants were trained to manipulate the documents, use the keyboard/table, and all the system features of IST. After learning all of the features, the participants were given 5 minutes to practice using the features, walking around in the headset, and using the keyboard/table.

After completing the tutorial, the participants were given a quick break where they could take the headset off. During this time, they were provided a printout version of the experimental task, as seen in section 4.2.3, and the participants were encouraged to ask questions if clarification was needed before beginning the study. Once the participant was ready, they would don the headset and begin the task.

Session - Participants were instructed to start their analysis and work towards an initial hypothesis. Participants were given about 40 minutes for this session. Participants were encouraged to ask questions on the task or the system if they needed help.

Interview - Session 1 concluded with a 5-minute post-session interview to gather information on the strategies and organizational structures used during the analysis task thus far.

Session 2

Onboarding - Before beginning the study portion of the session, the participants were re-familiarized with how to use the equipment and all the adjustments on the headset.

Session - The participants were given 65 minutes to continue their analysis and develop an outline of their findings. If, at the 50-minute mark, the participant had not started their outline, they were instructed to do so.

Interview - Session 2 concluded with a 5-minute post-session interview aimed at collecting information on the strategies and organizational structures used. Additionally, we wanted to understand if the structures/strategies used changed from the previous session.

Session 3

Session - The participants were given 60 minutes to finish the analysis and develop their 1-2 page report of their findings.

Interview - During this 15-minute interview, the participants answered questions about

the task, including overall strategy, overall spatial layout, and how the layouts changed over time.

During each session, the participants were given time warnings at the halfway point and 15, 5, and 1 minute remaining. This was to help the participants use their time efficiently and keep track of their progress.

4.3 Results and Discussion

To evaluate RQ 1.5, we looked at the overall structures formed by the participants during their analysis and how those structures changed across their analysis process. For RQ 1.6, we looked at mapping the 3D layouts and the 1D reports written during the analysis using the document reference (citation) order. For RQ 1.7, we focused on the participant's movement in the physical space during the analysis task. Lastly, for RQ 1.8, we looked at the system-level interactions and overall analysis strategies to understand the relationships between actions and report scores.

4.3.1 RQ 1.5: Spatial Layouts and Change Over Time

Spatial Layouts

To understand the overall spatial structure formed by the analysts, we use a bird's eye view into the analysis space to understand the high-level organizational pattern used for document layout. In previous work, four high-level organizational patterns have been detected within immersive spaces. Using the updated definitions of these organizations proposed by Davidson et al. [20], these layouts are *Semi-Cylindrical, Cylindrical, Environmental, and Planar*. Semi-cylindrical is defined as having documents curved around the user but not on all sides as seen in figure 4.3 left, whereas cylindrical layouts have document positions in the full 360 degrees around the user as seen in figure 4.3 middle. Environmental Layouts use the physical features of the space, such as the floor/bulletin board for organizing the documents onto a plane as seen in figure 4.3 right, and planar layouts use planes within the space that are not curved around the user while also not matching a physical feature of the space as seen in environmental.

To categorize the overall spatial organization of their analysis space, we look at the final layout created by the participants. This study found four semi-cylindrical layouts, six cylindrical layouts, and one environmental layout.

Organizational Features Digging deeper into the overall structures formed, we can also look at what additional features the participants used to create organization within the

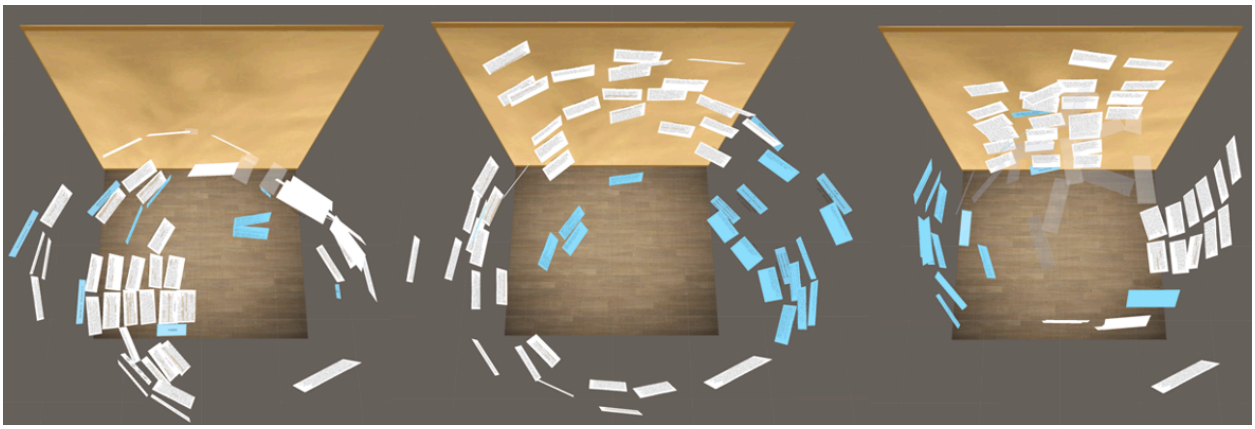


Figure 4.3: Top-Down View at High-Level Organizational Layout of the analysis spaces. Left: Semi-Cylindrical Layout, Middle: Cylindrical Layout, Right: Environmental Layout. The white documents are those provided for analysis, and the blue documents indicate a note or label created during the analysis. In the bottom right-hand corner is a copy of the task as detailed in section 4.2.3.

spaces. Previous work by Davidson et al. examined five other deeper organizational elements within the final layouts of the participants, including relevance vs distractor, plotline encoding, reporting agency sorting, timeline formation, and trash piles. We examine these features within the space below and two new features not yet found within the final layouts - network graphs and scratch space.

Relevant vs. Distractor Documents

One feature of the space evaluated in the past was the distance of relevant or distractor documents to the center of the tracked area. In the analysis of the professional users, we found that distractor documents had an average distance of 2.256 m (SD = 0.6898 m) to the center of the tracked space, with relevant documents having an average of 2.290 m (SD = 0.6842 m) to the center of the tracked space. At first glance, it seems concerning that the relevant documents are further from the center of the space. However, when looking at the overall organization of these spaces, the participants' schemas (document locations) are not often centered within the tracked space. For this reason, we chose to calculate the center point of the space based on where the participant spent the most time (dwelled) during their analysis. Then, using this new center point, we calculated the relevant and distractor document distances to that point. In this new calculation, we found the distractor documents were, on average, 1.428 m (SD = 0.3283 m) from the center, and relevant documents were 1.388 m (SD = 0.3047) from the center. Using a t-test, we did not find that these were statistically significant ($p = 0.1825$), but we suggest that future work verify these results with large participant pools.

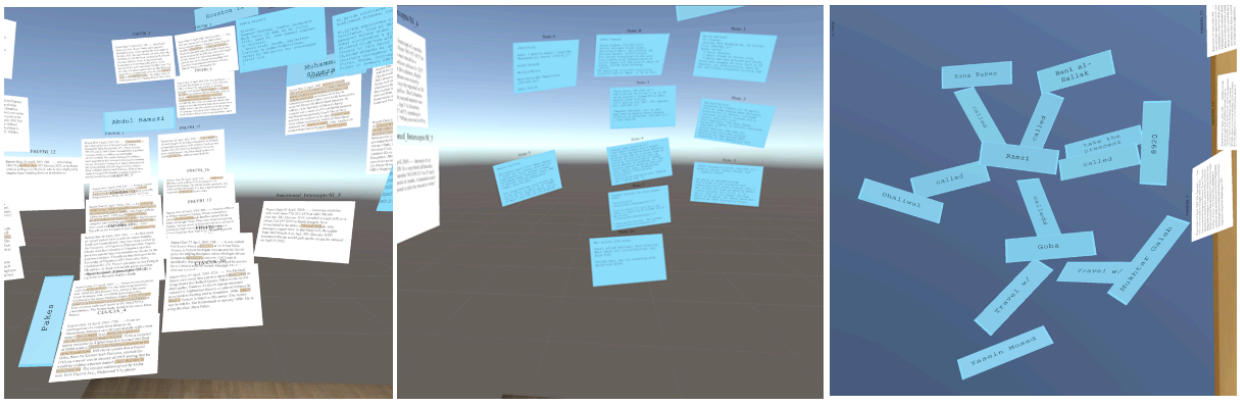


Figure 4.4: Deeper Organizational Features found in the analysis spaces. Left: Column/Timeline organization with additional labeling of people of interest within the columns, Middle: Scratch Space for the analyst's comments, Right: Network Graph created with the label feature.

Reporting Agency Sorting

Another feature we found within the deeper organization strategies was sorting by reporting agency according to the task dataset 4.2.4. In this study, we had four participants who organized the data into distinct categories based on the reporting agency of the documents. Figure 4.3, right has an example of the reporting agency sorting where the FBI reporting documents are organized on the bulletin board, the CIA documents are organized on the right-hand side, and the Sanctioned Intercepts documents on the side of the space opposite from the bulletin board.

Timelines

In our analysis of the spatial layout created by the participants, we identified that all participants (11/11) utilized a timeline column during their analysis of the documented set. Figure 4.4 Left shows an example of two long-column timelines.

Scratch Space

Lastly, in our analysis of the deeper organizations used within the space, we also identified the use of scratch space with three of the participants. An example of scratch space can be seen in figure 4.4 middle. In scratch space, a designated section of the analysis space was used to organize the analyst's notes/comments, separate from the provided text documents. This approach to managing their notes created a division between the ground truth of the dataset (i.e., the provided text documents) and the analyst's thoughts/comments (i.e., interpretation

of the ground truth provided).

Network Graphs

A new feature we discovered in analyzing these spatial layouts was a network graph generated by P10, as seen in figure 4.4. In their analysis, they used the ceiling of the analysis space and the label feature to create a social network map. For example, the participant had a "Person 1" label with another label, "Called", then another label, "Person 2," allowing them to build a graph of Person 1 → Called → Person 2 →, etc., making the information into a more coherent narrative and glanceable. While we only had one user create a social network graph this way, we want to highlight the use of this deeper organization created by the participant and suggest building support for generating graphs like these in the future.

Change Over Time

To understand how IST supports different stages of the sensemaking task, we wanted to understand the following:

1. How do the spatial layouts change across the analysis task?
2. How does document interaction change across the analysis task?
3. How do these trends compare to those found in a previous study using novice analysts?

To understand how IST supports the overall sensemaking task, we look at how the spatial layouts created by the participants change across the analysis sessions. In doing so, we can look at the evolution of the schemas formed during the sensemaking process. An example of how P5's spatial structures evolved across the analysis can be seen in figure 4.1. In this stepped evolution, we can see a semi-cylindrical layout begin to form by the end of session 1, all of the documents being removed from the bulletin board by the end of session 2, and some columns/timelines forming on the back side of the tracked space into the end of session 3. In addition to visually inspecting the spaces and how they changed across the sensemaking task, we wanted to evaluate how the users grouped information over the sessions and if those groups became more distinct (i.e., more organized). For this analysis, we leverage clustering algorithms as seen in previous work to understand how the overall organization changed over time [20]. For our analysis, we ran OPTICS (Ordering Points to Identify the Clustering Structure) clustering [3] on the final save of each session created by the participants. We chose to use the OPTICS clustering technique because it performed well at clustering the columns and clusters formed by the participants when we compared it to K-Means [67], Hierarchical Clustering Techniques [69], and DBSCAN [30]. In addition, OPTICS is a density-based clustering technique that also detects noise (documents too far

away from others to be considered part of a cluster) within the algorithm. In our OPTICS analysis, we used min samples=2, metric=l1 with the remainder of the parameters set to the defaults within [72]). An example of the clustering produced by OPTICS can be seen in figure 4.5

To evaluate the clustering over time, we utilized two different clustering evaluation methods - Silhouette Scores [76] and Calinski Harabasz (CH) Scores [13]. Silhouette scores range from -1.0 to 1.0 and indicate how close a point or document is to its own cluster (cohesion) compared to the other clusters (separation). CH is also known as the variance ratio criterion and is an indicator for the sum of between-cluster variance and within-cluster variance for each cluster. With CH, the higher the score, the better the overall clustering, and for silhouette scores, the closer to 1.0 indicates a very *strong* clustering. Overall, we expected to see that the silhouette and CH scores would increase across the sensemaking task.

The average silhouette scores were 0.32, 0.40, and 0.42 for sessions 1, 2, and 3, respectively. We performed an Analysis of Variance (ANOVA) test with a random effect of participant to account for individual differences between participants, using Tukey's HSD method for post-hoc comparisons which adjusts for multiple comparisons and controls the overall Type I error rate. We found a significant difference in silhouette scores ($F(2,20) = 4.8452$, ($p = 0.0192$)). The post-hoc tests revealed a statistically significant difference between session 1 and 3 ($p = 0.0214$) with session 3 higher than session 1, and a trend towards significance between sessions 1 and 2 with session 2 higher than session 1 ($p = 0.0690$). There was no difference between sessions 2 and 3 ($p = 0.8403$).

The average CH Scores were 43.03, 54.55, and 56.49 for sessions 1, 2, and 3, respectively. We found a significant difference in CH scores ($F(2, 20) = 5.025$ ($p = 0.0171$)). The post-hoc tests revealed a statistically significant difference between session 1 and 3 ($p = 0.0215$) with 3 higher than session 1, and a trend towards significance between session 1 and 2 ($p = 0.0522$) with session 2 higher than session 1. There was no difference between sessions 2 and 3 ($p = 0.9060$). Combining the overall change over time graphics as seen in figure 4.1 and the clustering analysis, we can see that these spaces are 1) constantly changing, 2) becoming more organized during the analysis, and 3) most interestingly, the spaces (spatial structures) are refined up to the final stages of sensemaking where the participant is preparing the presentation of their findings in the written report.

Next, we were interested in understanding the overall trends of the document movement across the sensemaking sessions. Based on previous analysis, we expected to see a decrease in the distance a document moved across the session and the number of documents moved in each session. The analysis of document movement trends over each session can be seen in figure 4.6a and 4.6b. Overall, we saw a decrease in the number of documents moved per session from sessions 1-3 and 2-3. We also saw that the distances those documents moved decreased from sessions 1-3. Using an Each Pair Students T-Test with Bonferroni's correction, we saw no statistical differences between distance or count moved from sessions 1-2.

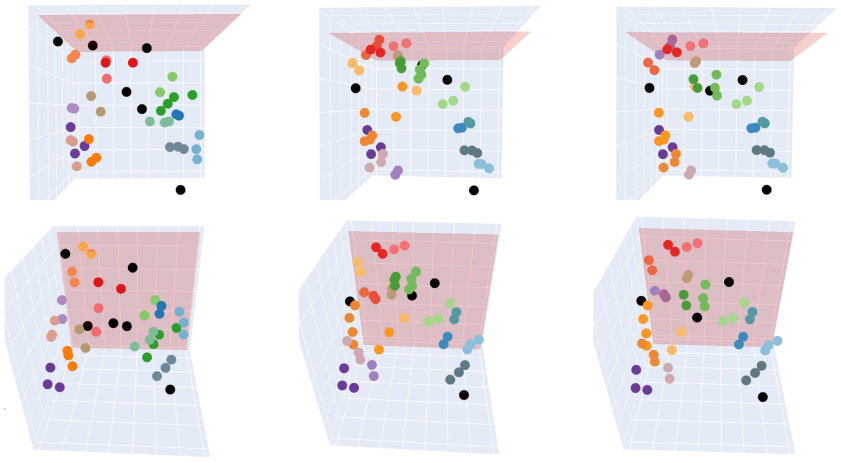


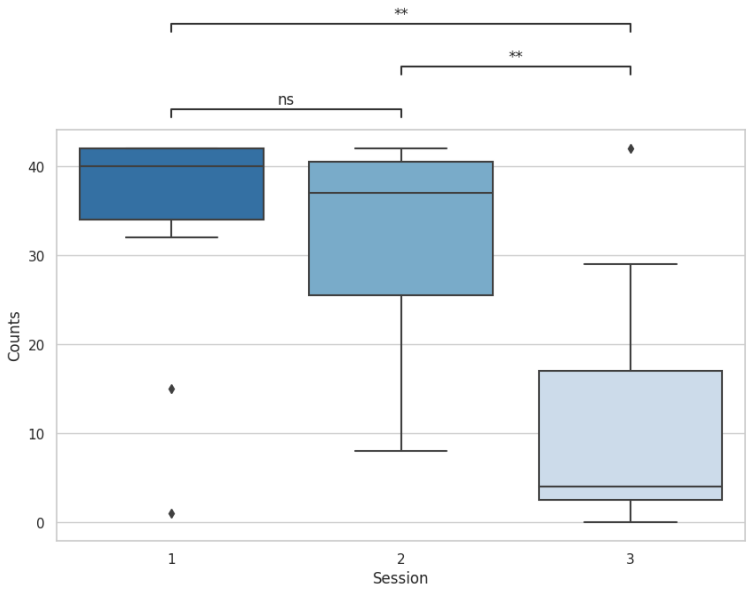
Figure 4.5: From left to right, End of Session One, End of Session Two, and End of Session Three of P4's analysis space. The top row is a top-down view into the space, and the bottom row is a rotated view looking towards the bulletin board. The colors in this image indicate the clustering applied to each document by the OPTICS algorithm. The black documents were "noise" detected by the OPTICS algorithm.

Novice vs. Professional Analysts

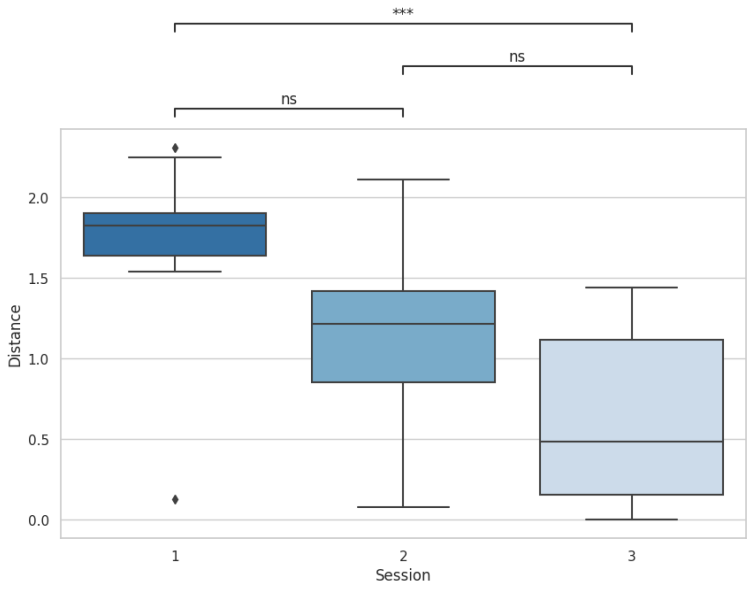
Finally, to address RQ 1.5C, we look at how the professional analyst populations' overall trends compare to the novice analysts' identified trends in previous work. First, in high-level spatial structure analysis, we see that the analysts' population structures match those identified in previous work. The analysts formed semi-cylindrical, cylindrical, and environmental layouts within the immersive space. In addition to seeing trends between the populations with the high-level structures used for document organization, we also found some trends within the deeper organizational features of the immersive spaces. One feature we saw every analyst use was timelines/columns, which was also seen with 6/8 novice users.

A feature we saw with only one novice, sorting by reporting agency, was identified in 4/11 analyst sensemaking layouts. This work also uncovered two new organizational features that were not identified with the novice population: Network Graphs (1/11) and Scratch Space (3/11). While these deeper organizational features are not as popular as the column/timeline pattern, we believe they are essential to support future IST systems designs because they give the analyst more control over the organization or location of details within their space.

Additionally, our analysis of the professional analyst population's spatial structures identified similar patterns for how the spaces evolved through incremental refinement. Overall, most of the documents are moved in analysis early on in the sensemaking process (sessions 1 and 2) with a decrease in the count but not zero movement into the later stages (session 3). In addition, we see a decrease in the distances that these documents move from session 1 to 3 with less overall distance as the analyst moves further into the sensemaking process.



(a) Number of documents moved per session by each participant.



(b) Average distances the participant moved a document during each session.

Figure 4.6: The total number of documents moved was significantly different from session 1 to 3 ($p = 0.004$) and again from session 2 to 3 ($p = 0.007$). There are also significant differences in the average distance a document moved from sessions 1 to 3 ($p = 0.0007$) using an Each Pair Students T-Test with Bonferroni's correction.

In summary, we found that the analyst population utilized the same high-level structures previously identified within IST analysis. This confirms that IST systems may benefit from previously proposed assistive high-level organizational features. Additionally, the analyst population utilized the same deeper organization features previously identified in IST analysis, and some new organization features not yet identified. Using these findings, we can verify many of the previous results from IST systems with professional users and provide new insights into the features that can be developed to better support the overall sensemaking process in systems like IST.

4.3.2 RQ 1.6: Mapping of Spatial layouts to Report Structure

Along with an in-depth analysis of the spatial structures and organization of the analysis space, we were also interested in understanding how the analysis space mapped to the reports written by the participants. The objective of this analysis was to evaluate the potential of future assistive report-generating features (e.g., automatically generating a report outline based on the spatial layout of documents in IST). In the instructions given to the participants, they were asked to include references to reports when citing information from them. Then, using these references, we generated a reference order of the reports (e.g., CIA_8, FBI_2,....SI_4). Then, using this ordered list of references, we located the document within the immersive space in the 3D scatter plots and drew a line through the immersive document layout, as we can see in figure 4.7. This line represents the documents directly referenced within the written report, which was then colored based on which paragraph the document was cited in (starting with red and ending in pink).

We drew this line on each final layout and then calculated the average total distance along the reference path of 11.25m (sd = 5.0974) for each participant. This total distance represents the distance between the first point to the second point, the third point, etc. Additionally, we looked at the distance along the reference path normalized by the total pairwise distance between all documents within the space (spread), which had an average of 0.00823m (sd = 0.0045). This value gives us a better understanding of the total distance along the path relative to the total spread of documents within their spatial layout.

The next feature of the reference lines that we evaluated was the shape of the pathway through the space. In previous work, three pathways were presented: left $\rightarrow$ right, right $\rightarrow$ left, and no clear path, meaning there were many intersections along the line with no clear shape to the line. For clarity, we will rename these to clockwise (left $\rightarrow$ right) and counterclockwise (right $\rightarrow$ left). An excellent example of a clockwise reference pattern can be seen in figure 4.7 left. The red documents are all cited within the first paragraph, and then we can follow the reference order in a clockwise pattern through to the pink documents, which are cited in the last paragraph together. In our analysis of the professional analysts, we found five roughly clockwise patterns, three roughly counterclockwise patterns, one new pattern, which we call top-down, and two participants with no distinct pattern. In the top-

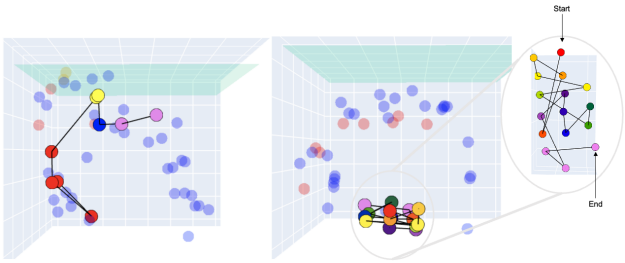


Figure 4.7: Left: P3 spatial layout with a reference line drawn through it. The green mesh indicates the bulletin board. The red, opaque documents are cited in paragraph one, the yellow documents in paragraph 2, the blue in paragraph 3, and the pink in paragraph 4 of the report. Right: P5 reference order of moving relatively top-down through the columns within their analysis space for citing documents. The zoomed portion is oriented as if the participant is standing in the center of the analysis space, staring at documents. Most Red, Orange, and Yellow documents are on the top of the columns, with Pinks, and Purples and Blues toward the bottom of the columns. In this figure, the documents are cited in paragraph order Red, Orange, Yellow, Green, Blue, Indigo, Violet, Pink (ROYGBIV + Pink).

down pattern, the participant referenced documents at the top of columns created in their analysis first, then they worked their way down the columns toward the bottom. Figure 4.7 right shows an example of the top-down layouts.

Novice vs. Professional Analysts

In looking at the document reference paths in the analysis space, the novice reference paths were longer than those of the professional analysts. However, this was due to the total tracked analysis space of the novices being a 4x8 meter tracked space compared to the 3x3 meter tracked space provided for professional analysts. When we look at the total normalized values of the reference line, we do not see any statistical significance between the professional and novice user populations. However, when we look at the reference line's overall shape, we see more professionals with a roughly identifiable transformation path than the novices. In the shape analysis, we found that 9/11 of the professional analysts had a clockwise, counter-clockwise, or top-to-bottom transformation path compared to the novices, where only 4/8 had a roughly clockwise or counter-clockwise path. This could be due to the professional analysts having a better organization overall than the novice users or a better ability to cluster relevant information for the report writing.

4.3.3 RQ 1.7: Physical Navigation

One of the study's goals was to understand how the analysts navigate the immersive space (i.e., walking) while they complete the sensemaking task. To evaluate the participants'

physical navigation, we tracked their movement during their analysis to understand the space usage while they were completing their sensemaking. In our files, we collected the headset position x, y, z, yaw, pitch, and roll ten times per second while the participant was in the headset. We took the average headset position, once per second, to use in the following analysis.

Professional Analysts Physical Navigation

Across the entire analysis task, we found that the participants navigated 434.32 m (sd = 138.64) of the space. With breakdowns of 131.04 m (sd = 36.38 m) in session 1, 201.78 m (sd = 75.92 m) in session 2, and 101.49 m (50.04 m) in session 3. To account for variation in the amount of time spent in the headset, we also present the amount navigated normalized by number of seconds spent in the headset: total average normalized 0.11 (sd = 0.050), session 1 normalized 0.05 (sd = 0.02), session two normalized 0.04 (sd = 0.015), and session three normalized 0.03 (sd = 0.009). Additionally, we visualize this data in figure 4.8.

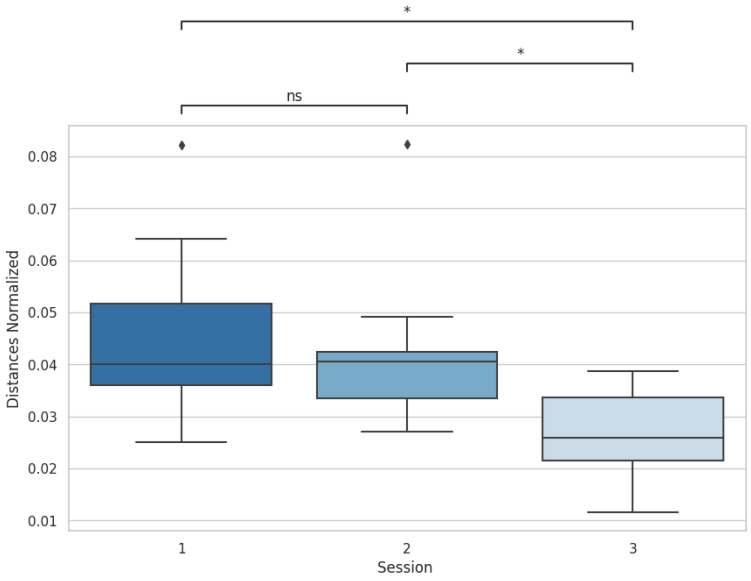


Figure 4.8: Movement data normalized by time in the headset. We can see a decrease in movement over time with significant differences between sessions 1 and 3 ($p = 0.02$) and 2 and 3 ($p = 0.03$) using an Each Pair Students T-Test with Bonferroni's correction.

When we look at the data normalized by the amount of time spent in the headset, we can see a downward trend in navigation as the participants work into the later stages of the sensemaking process. We see no statistically significant difference in the amount navigated between sessions one and two. However, there was a difference in the amount navigated between sessions 2 to 3 ($p < 0.05$) and sessions 1 to 3 ($p < 0.05$) using Each pairs Student's T-Tests for pairwise comparisons with Bonferroni's correction.

Considering the task provided, it makes sense that the participants spent less time navigating the space during session three, which focused on report writing. However, when we look at the average navigation (pre-normalization), we can see 101.4930 m on average were navigated by our participants during that session, showing that, even in the late stages of the sensemaking process, using physical navigation to access externalized memory to be used by our participants.

In addition to looking across the participants for overall trends in movement, we also looked deeply at each participant's movement trajectory as seen in figure 4.9 and their dwelled areas as seen in figure 4.10. One trend we found when we looked deeper at the individual participants was two groups of participants when it came to navigation within the immersive space. We had participants who navigated more of the space - **Movers**, and then we had participants who were primarily stationary during their analysis - **Stationary**. This is supported by the large standard deviation we found in the above data. In figure 4.9, we can see a mover on the top row compared to a stationary participant on the bottom row. Overall, we had six movers and five stationary participants. In our analysis, we looked at both the shape of the movement trajectory and the amount of space navigated by the participant.

In addition to looking at the overall trajectory of the movement, we also looked at where the participants spent time during their analysis, which we define as dwell. When we look at the overall trends of dwellings, we see a decrease in the number of intense dwelled locations across the sessions, indicated by a darker green color in the heatmaps. For instance, with P6 in figure 4.10, we can see the number of locations with intense green (more time spent in that location) decrease across the session to one primary dwell location during the final session. While this trend is more pronounced in the mover group of participants, we also saw this trend across the stationary participants.

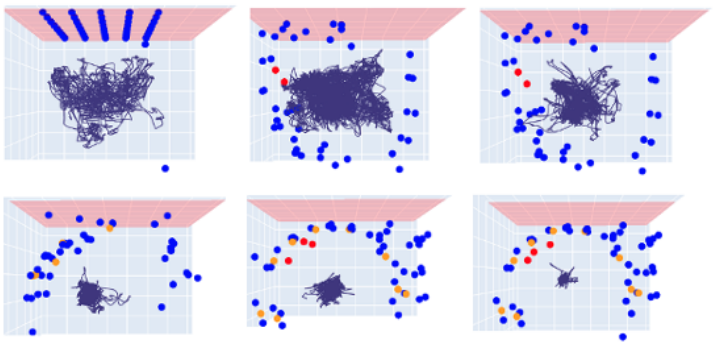


Figure 4.9: From left to right, session 1, session 2, and session 3 for P6's (top) and P11's (bottom) analysis. The red plane represents the bulletin board, the blue points indicate documents, the red colored points are notes created by the user, and the yellow points are labels created by the user.

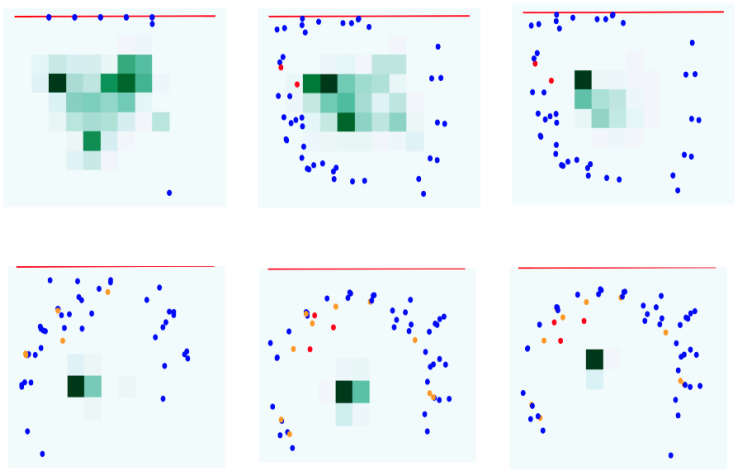


Figure 4.10: From left to right, session 1, session 2, and session 3 for P6's (top) and P11's (bottom) analysis. The red plane represents the bulletin board, the blue points indicate documents, the red colored points are notes created by the user, and the yellow points are labels created by the user. The heat maps show the time spent (dwell) within a one sq ft area of the tracked space.

Novice vs. Professional Analysts

Overall, the patterns we saw with physical navigation were similar to those of our novice users.

In our analysis of the professional analyst movement in IST, we saw a couple of trends. First, we saw decreased movement as the participants moved further in the sensemaking process. Additionally, we saw two trends in movement style in the space - **Movers vs Stationary** users. Lastly, we saw trends in the decreased number of intense dwell spots over time, as seen in our analysis of time spent in the different regions of the analysis space.

When we compare these findings to those reported in chapter 3, we see many similarities between novice and professional analysts. When looking at the normalized movement data, we see a decrease in movement from sessions 2 to 3, which is supported in both the analyst and novice populations. However, with the professional analysts, we also saw a decrease in movement from sessions 1 to 3, which we did not see in our novice population.

Another similarity we found was with the patterns in the trajectory of movement between the analysts and the novices. In both cases, there was a distinction between the movers who navigated most of the space while conducting their analysis and the stationary who tended to stay in one spot. We also found similar trends in the dwell analysis that looked at time spent within square foot regions of the space. Overall, we saw a decrease in the number of dwelled locations as the participants progressed further into the sensemaking process, with the caveat that this trend is more evident within the **mover** participant group than with

the stationary group.

Overall, the analysis of movement data between the professional and novice analysts confirmed the findings in chapter 3. Overall, the participants tend to decrease movement as they advance into the later stages of the sensemaking process. There are two types of analysis styles when it comes to movement within the IST system, and we see a decrease in the number of dwelled areas, more obviously in the movers; however, we do see similar trends in the more stationary participants.

4.3.4 RQ 1.8: Strategies and Report Scores

Our last research question, RQ1.8, aimed to understand if specific spatial layouts, strategies, or interactions within IST correlate to better quality sensemaking. We needed an indicator for sensemaking quality for our analysis, so we relied on the reports written by the analysts during session three of their analysis. We evaluated these reports for two criteria - Correctness and Quality using the rubrics described in section 4.2.6.

Correctness - To grade the overall correctness of the reports, the experimenter evaluated each report in a randomized order. A point was awarded for each correct piece of information provided in the report based on the ground truth of the dataset.

Quality - To grade the reports for overall quality, we worked with a set of professional intelligence analysts. Before grading the reports, the experimenter met with the analysts to onboard them to the grading process and provide them with detailed instruction sheets including a document that specified the order to grade the reports (randomized for each grader). The graders were then provided a PDF of each of the reports written by the participants, and the grading rubric via a Google Form. To ensure the graders were provided the same materials as those who graded the novice participants' reports, the analysts who graded these reports were also provided with *Getting Started* materials that included the task the participants were given 4.2.3, the solution to the analysis, and the documents used during the analysis, which they were able to reference if needed during the grading process. The analysts were given a month to find flexibility in their work lives to complete their grading and asked to complete their grading within 1-week after starting. In total, we had three professional analysts complete the quality report grading.

High-Level Strategies

First, we take a high-level look at the overall strategies that the professional analysts utilized for their sensemaking within IST. For these high-level categorizations, we build on the previous work of Davidson et al., and Kang et al. [20, 48] In this study, we identified three high-level strategies for the analysts' sensemaking. "Build from Detail" (S1 in Davidson et al.), where the participants read all documents first, then sort the documents into groupings,

and refine those groupings as needed. Overall, we have two participants who employ this sensemaking strategy. Next, we have "Hit the Keywords" (S3 in Davidson et al.), where the participants skim documents for keywords, execute searches on those keywords, and then group and refine based on those queries. We had 6 participants employ this sensemaking strategy. Next, we had "Overview, Filter, and Details" (S4 in Davidson et al.), where the participants get an overview of the content of the documents first, filter or sort the documents into rough groupings (normally based on reporting agency information) and then revisited relevant documents for details and deeper understanding. We had three participants employ this method.

In our analysis of these strategies, we attempted to correlate the strategies and overall sensemaking performance. We did not find any statistically significant correlation between the overall sensemaking strategy used and the report's correctness or quality scores scores.

Note Taking

Another item we were interested in was the use of notes during the analysis process. In our comparison of the two populations, we found that the total notes created by analysts (mean = 5.09, SD = 4.13) were more than that of the novice analysts (mean = 2.63, SD = 1.41) using a t-test (p=0.0446). However, the total number of notes does not indicate the quality or content of the notes.

Therefore, to dig a bit deeper into this, we were interested in how we can evaluate the content of the notes to understand if there are more profound differences in the types of notes used during this analysis. To evaluate the quality of the content of the notes, we looked at the total number of notations (all marks made in the notes), the total number of words (words, symbols, and abbreviations of the notes), and lastly, the total number of content words (nouns, verbs, adjectives, and adverbs) identified by Siegel [83]. For this analysis, we used the following steps: 1) get all notes (analyst and novice), 2) drop any empty notes, 3) count the number of notes for each participant, 4) combine all of the notes' content into a text file, 5) create a cleaned version of note content, 5) count the total number of characters in note content, 6) split the note content into words, 8) count the words, 9) apply the Natural Language ToolKit's (NLTK) part-of-speech tagging to each word in the notes [9], 10) count the total number of parts of speech, and 11) normalize each part of speech by total word count.

This analysis found two interesting differences between the novice and professional analysts' notes. The first difference we detected was that the professional analysts (mean = 130.21, SD=63.73) used more nouns than the novices (mean = 72, SD= 93.10) (p=0.0406) and the professionals (mean = 11.77, SD = 7.78) utilized more digits than the novices (mean = 2.73, SD = 4.38) (p = 0.0045). This finding was quite interesting since we did not find any difference in the notes' total number of characters or words. Indicating the analyst notes contained more pertinent content relating to people, places, things, phone numbers,

addresses, and bank account information than the novices.

Overall Highest vs Lowest Scorers interactions

Taking our analysis further, we wanted to understand trends/differences between the overall highest and lowest scorers. For this analysis, we combined the data from the professional intelligence analyst and the novice users. Then, we took the five highest (4 Analyst, 1 Novice) and five lowest (1 Analyst, 4 Novice) scores from the combined population. Using the combined data, we created ridge line plots looking at the interaction frequency of each log file interaction (e.g., highlights, searches, grabs, etc.) between the two groups. We normalized each interaction between 0 and 1 for comparison purposes, and the resulting ridgeline plot can be found in figure 4.11.

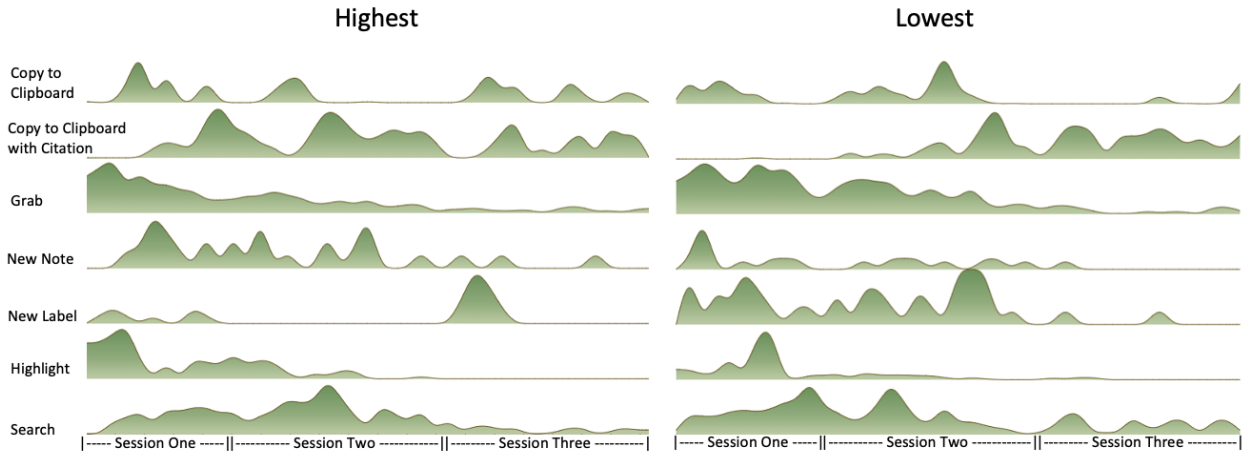


Figure 4.11: Left - Highest scorers, Right - Lowest Scorers. The x-axis represents time within each of the sessions from the start of the analysis to the end of the analysis. The y-axis is the frequency of each interaction normalized between 0 and 1.

One interesting difference we found was that the highest scorers tended to use the copy-to-clipboard features earlier in the sensemaking process than the lower scorers. Another interesting difference between the highest and lowest scorers was using notes. The highest scorers tended to create new notes more frequently than the lower score during sensemaking. Interestingly, however, lower scorers tended to produce more labels than the higher scorers. Another interesting difference we detected between our highest and lowest scorers was the frequency of searches during the third analysis session. The lower scorers tended to rely on the search feature more during this last analysis session, which could indicate that they relied on the software to locate critical information they planned to use in the report writing versus relying on their externalized memory to access that information.

System Log Interactions

Another feature of the analysis we were interested in was whether specific interactions with the system during the analysis (e.g., highlighting, searching, note-taking, etc.) correlated with higher or lower performance on the sensemaking task. The idea is to learn from what the professional analysts did during their sensemaking to guide what leads to effective sensemaking within systems like IST. In this analysis, we found two statistically significant correlations. First, we found that the total number of notes positively correlated (0.5197) with the report's correctness (p=0.0259). Additionally, we found that the number of notes created during session one positively correlated (0.5640) with report correctness (p=0.0188).

Novice vs Professional Analysts

We found similar trends when comparing the professionals to the novice analyst populations studied before. When we look at the high-level strategies, we identify very similar high-level strategies to those found with the novice users. In our deeper analysis of feature usage, we found that the professional analysts produced more notes with more relevant content during their sensemaking than the novice users.

In addition to the high-level strategies, we looked deeper at the system-level interactions through the log file interactions. In this analysis, we found differences in the quality of the notes between the two populations and correlations between note creation and the correctness score. Lastly, we presented some trends in interaction frequency between the highest and lowest scorers across the two populations. Including search frequency late into the sensemaking process, note creation, label making, and copying to the clipboard.

4.3.5 Key Takeaways

At a high level, our analysis identified many similarities between the novice and professional user populations. There are several possible reasons for these similarities. One interpretation is that the sensemaking affordances and features of IA systems, such as IST, enable novice users to perform similarly to professionals in sensemaking tasks. Perhaps the ability to externalize the analysis process by organizing documents and annotations in 3D space helps novices think more systematically and, therefore, perform more like a trained analyst. Another perspective is that when professional users engage in an analysis task using an unfamiliar system like IST, they become novices within that system despite their professional sensemaking skills. This speculation is based on the familiarity of the tool to the participants. For instance, novice users have no expectations about how intelligence analysis tools should function, while professionals work with specific tools daily, expecting them to support their analysis process optimally. Therefore, when presented with IST, professional users are asked to perform an analysis task without their familiar tools.

Similarly, another interpretation regarding our results could be rooted in the analyst's ability to apply their expertise/domain knowledge during this task. In intelligence analysis, analysts often focus on specific areas, building domain expertise, similar to a student concentrating research efforts on a particular topic (e.g., Immersive Analytics) during their dissertation. However, in our study, professional analysts complete a fictional analysis task where their specific domain expertise cannot be applied. This suggests that we removed some domain-specific elements from the professional analysts' analyses, leading them to perform more similarly to novices within the system. While all three interpretations likely have some merit, since the overall correctness scores were low, we suggest that the second and third interpretations are most likely. In other words, we speculate that IST does improve the analysis performance of novices [59] but also reduces the analysis performance of professionals, at least initially.

In our study, we observed similarities between the professional and novice users. However, we also noticed some intriguing differences between these populations. One possible interpretation of these differences is that, while the professional users may have experienced some reduction in familiarity and domain-specific knowledge, these professionals are skilled in sensemaking and the analysis process, which could have contributed to identifying these interesting and novel differences, such as the analyst who built a network graph or use of the utilization scratch space to keep *analysts comments* separate from the ground truth of the dataset.

Overall, we believe our results suggest that novice user populations can provide our research community with novel insights into immersive analytic sensemaking in traditional user studies. However, to gather these insights during an inherently challenging task such as text-based sensemaking, it is essential to identify highly motivated novices, i.e., those interested in the analysis task. This can be facilitated by providing user-study payment [19] or recruiting participants interested in the experimental task subject[60, 61]. Additionally, we believe that our results could suggest a need for more genuine longitudinal user studies with professional users, which could allow us to see more significant differences between user study populations due to the longitudinal nature, allowing the professionals to get over the unfamiliarity of using the system during analysis. However, there are many logistical difficulties in running a genuinely longitudinal study. Lastly, in a similar vein, we believe that in other non-traditional user studies (e.g., expert feedback sessions or in-the-wild case studies), it is essential to gather feedback from professional users to gain insights into how a prototype can better support them in their daily tasks.

4.4 Limitations

We recruited 11 professional analysts with an average of 9.72 years of analysis experience. However, this is still a small sample size regarding power in statistical analysis. Due to this, we report all of our findings with caution that future work should be done to help validate

the results.

Additionally, to recruit professional intelligence analysts in a convenient way, we set up the experiment in a room accessible to the analysts on the campus where they worked. Unfortunately, this room only had about 3x3 meters of open space for tracking, which led to a varied amount of tracked space between the professionals and novice populations. Where necessary in this paper, normalization has been applied to draw comparisons between the groupings. We do not believe this has affected the results of this study but still present it as a possible limitation.

Another limitation of this work was surrounding the text-entry system used to follow the user-study design of Davidson et al.'s prior work [19]. A slightly inaccurate keyboard model was placed on the VR environment table, to help participants locate the keyboard on the tracked table. This may have caused challenges for participants when entering text within the system. Future work in immersive analytics could utilize AR or AR pass-through portals for text entry, as seen in Giovannelli et al.'s work[35].

4.5 Conclusions

This chapter aimed to fill a gap in the existing literature by providing an in-depth evaluation of professional analyst sensemaking in an immersive analytic system. Our analysis found that the high-level organizational structures used by the analysts were similar to those discovered using novice user populations. Additionally, we found that the professional analysts utilized all of the previously discovered deeper organizational patterns presented in chapter 3, adding two new features: Network Graphs and Scratch Space. Additionally, we found that the overall movement patterns of the professional analysts matched those of the novice users. Lastly, we identified new trends between the highest and lowest scores on the reports produced at the end of the sensemaking task.

Using these findings, we speculate on why we identified many similarities within the user study populations and suggest that for future work in more traditional user studies, identifying motivated novice users can be sufficient in our community to gather key insights into user interactions within IA systems. We also suggest features that can be designed to support sensemaking in immersive analytic systems better. For instance, linking features can be added to allow users to externalize their cognition in new ways to help better the ability to create graphs. Additionally, when we look at the mapping of the 3D spatial layouts and the 1D reports of professional analysts, we have identified more distinct mapping patterns that show promise in creating assistance for report writing in this late-stage part of sensemaking. Future work could be done to examine how to transform the 3D spatial structure into a report outline using the information-rich data of document placement and generative large-language models. Additionally, more work can be done to examine how additional organizational features assist in the sensemaking process. Lastly, to help unify

the two sensemaking subloops, work on semantic interaction could be a fruitful next step in immersive analytic research.

Chapter 5

Uncovering Best Practices in Immersive Space to Think

In the previous chapters, we evaluated how users complete a multi-session analysis task in IST with both novice and professional users. Additionally, previous work in IST has demonstrated advantages for immersive sensemaking however, all of these works have focused on completing a single sensemaking task. An unexplored area lies in investigating how users adapt their analytical and organizational strategies when undertaking multiple analysis tasks within an IA system forming Research Question 2 (RQ2): How does familiarity affect initial strategies and layouts used for immersive sensemaking? To address this question, we conducted an exploratory study involving professional analysts to gain insights into how the use of these systems changes as users become more familiar with IA sensemaking.

5.1 Introduction

Immersive analytics is an emerging research field that combines visual analytics, data visualization, and virtual/augmented reality. It focuses on using immersive human-computer interfaces to support analytic reasoning and knowledge generation (sensemaking) processes [14, 84]. Sensemaking involves extracting information, generating hypotheses, and formulating findings to make sense of the world around us. It is crucial in various domains like journalism [15], research, and intelligence analysis [73].

Visual analytics research emerged to assist sensemaking by offering interactive visual interfaces [89]. As immersive technology becomes more accessible, immersive analytics systems have been developed to enhance sensemaking. While previous work has focused on understanding how immersive analytics systems can support sensemaking using qualitative or quantitative data, it fails to explore the persistence of initial strategies and spatial layouts across multiple sessions. Furthermore, it overlooks strategies' evolution and refinement as users become more familiar with the tool and its effective usage. Therefore, there is a gap in understanding how learning to use immersive analytics systems influences overall analysis strategies and whether refined strategies lead to more effective sensemaking. This work aims to fill this gap.

We conducted a multi-session user study where participants completed three separate sense-

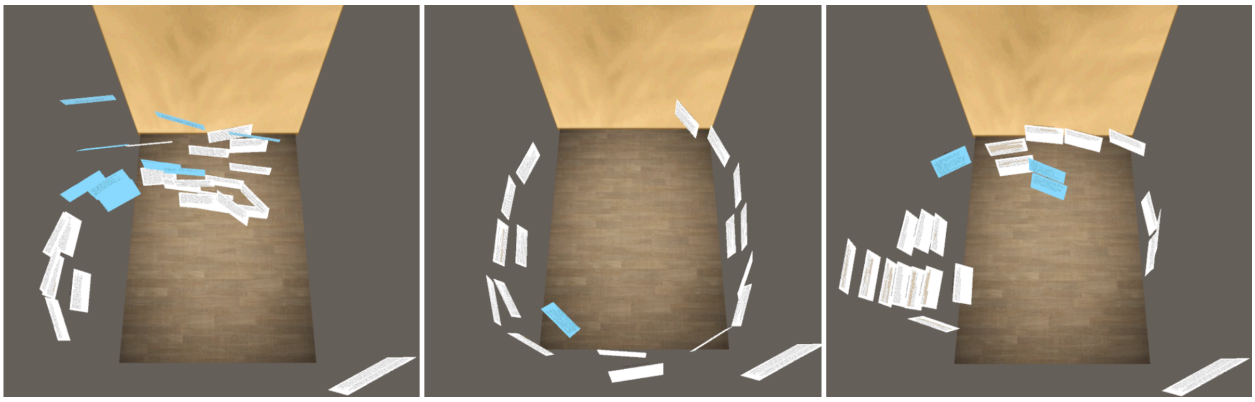


Figure 5.1: A top-down view into the final spatial layout of a participant. From left to right, session 1, session 2, and session 3. We can see distinct differences in the use of space and deeper organizational features used. The document located in the bottom right corner is the analysis prompt.

making tasks. The goal of this study was to provide users the ability to try different strategies for sensemaking as they became more familiar with the tool, as well as become more comfortable completing analysis in an immersive environment, in order to more deeply understand the evolution of their strategies and the effects of strategy change on task performance. We found that our participants used different spatial layouts over the different sensemaking sessions and utilized different strategies across the sessions. We also found correlations between certain patterns of user interaction and the quality of sensemaking analysis and that the overall organization of the sensemaking schemas increased over time.

The key contributions of this work are as follows. 1) An understanding of how learning and familiarity changes the sensemaking process in immersive analytics systems and how that influences the schemas and strategies of users. 2) Guidance on using immersive analytics systems for more effective analysis. 3) New data analysis methodologies for understanding sensemaking in immersive analytics prototypes.

5.1.1 Immersive Analytics System Usage and Learning

Immersive analytics systems provide many unique opportunities for data exploration, sensemaking, and decision making. However, in most traditional IA user studies, participants are trained on how to use a system and then are asked to complete a sensemaking task with that system [34, 59, 66]. While this provides initial knowledge on system usage, it does not allow participants to become familiar with the tools, try multiple strategies, and become comfortable completing sensemaking with them. As Norman pointed out, natural user interfaces are not always intuitive, and users may struggle with these interactions at first [70].

Additionally, biases such as anchoring bias [46] and mere exposure bias [97] may provide

challenges for users as they may feel more comfortable sticking to the first pieces of information presented to them or have a tendency to prefer familiar tools/strategies. When users are presented with an IA tool for the first time, they may struggle with interacting and understanding the new opportunities that these tools present. This work aims to understand how learning and familiarity affect sensemaking. Allowing participants to complete three separate sensemaking tasks on three separate days to become familiar with an IA prototype allows them to try multiple strategies and schemas and use different features within the prototype during their sensemaking task.

5.2 Immersive Space to Think Approach

The immersive analytics prototype used in this work builds upon the existing Immersive Space to Think (IST) approach [5, 19, 59, 60]. The IST approach uses a VR-based system, which provides a large-tracked area for completing a text-based sensemaking task. A top-down view of our prototype can be seen in Figure 5.1 with an example of how the documents can be organized during a sensemaking task. The text documents for sensemaking are loaded in a randomized order onto a virtual bulletin board. The document size is 0.5x0.3 (m), which supports about one paragraph of text (Text-Mesh Pro text size at 40) without the need for scrolling. During the sensemaking task, users can remove the documents from this bulletin board and place them anywhere within the 3D tracked space during their analysis. To support sensemaking, IST also provides users with a set of features to aid in their analysis. Our IST prototype supported *highlighting, searching, note taking, label making, and copying/pasting to/from clipboard*. For text entry, a tracked keyboard on top of a wheeled table was provided to our users.

Based on feedback from previous studies using IST, we added one additional feature was added to the IST prototype used in this study: **Quick Search**. This new feature enabled the ability to select a word directly in a document and search for it in the rest of the documents using a virtual button displayed on the side of the document. We still supported general search based on text entry, but the new quick search feature allowed our users to quickly see other instances of a word in the whole document set without needing to type the word manually.

5.3 User Study

5.3.1 Goals and Research Questions

This study aimed to understand how the learning effect of becoming more familiar with an immersive analytics tool affects the spatial layouts and strategies used during sensemaking

and how the refined strategies might lead to more effective sensemaking performance.

Our research questions were as follows:

Spatial Layouts

- ***RQ 2.1A*** What spatial layouts do users form in IST when performing sensemaking tasks?
- ***RQ 2.1B*** How do these change across multiple sensemaking sessions? (learning effect)

Strategies

- ***RQ 2.2A*** What strategies do analysts use in IST during sensemaking?
- ***RQ 2.2B*** How do these change across multiple sensemaking sessions? (learning effect)

Correlations

- ***RQ 2.3*** How do the spatial layouts, strategies, and interaction patterns in IST correlate to the quality of the sensemaking task?

5.3.2 Apparatus

For this study, we used an HTC VIVE Pro [1] VR head-worn display with a wireless adapter to allow for a tether-free VR experience. The HTC VIVE Pro has 1440 x 1600 pixels per eye (2880 x 1600 pixels combined), a refresh rate of 90 Hz, and a 110-degree field of view. A Steam VR 2.0 Lighthouse tracking system was used. The tracked area for the study was a 3x3 meter area in which the participants could freely walk around while completing their analysis. The participants were trained on the size of the space during the onboarding phase of the study, and the floor in the virtual space also represented the boundary of the tracked area. In addition, the experimenter watched to ensure that participants did not get too close to the physical room boundaries. For interacting with the system, participants used a Valve Index controller [2] in their dominant hand, and for text entry, a wheeled table with a Vive tracker [3] and wireless keyboard with number pad included were provided for the participant to move freely around the space as needed for their analysis. The 3D interaction technique used in this IST prototype was Ray-Casting with reeling. To assist the participants in typing, we cut out some of the foam padding from the bottom of the headset, which allowed participants to peek out of the headset to see their fingers on the keyboard. The IST system was developed and rendered using Unity Game Engine [4].

[1] www.vive.com/eu/product/vive-pro/
[2] valvesoftware.com/en/index/controllers
[3] business.vive.com/eu/product/vive-tracker
[4] unity.com

5.3.3 Datasets

We selected three separate datasets for sensemaking tasks. These datasets were presented to the participants in a randomized order to prevent unwanted effects of dataset order on the results of our study. The details of each dataset can be seen outlined below:

Sign of the Crescent is a fictional intelligence analysis training dataset comprising 40 text-based documents for analysis. This dataset has been used in previous studies [19, 58, 95]. In this study, we selected 20 documents from the original text corpus, with one main storyline, two locations of interest, and a few distractor documents. The documents were about a paragraph in length on average.

Manpads is a fictional intelligence analysis training dataset comprising 23 text-based documents for analysis. In this study, we selected 20 documents from the original text corpus with one main storyline, four locations of interest, and two distractor documents. The documents were about a paragraph in length on average. This dataset has been used in previous work [95] and had a similar format to that of Sign of the Crescent.

Stegosaurus is a 2006 VAST challenge dataset [39], which consists of 240 documents which are a mixture of news articles, maps, spreadsheets, and some other supporting documents. However, only ten documents have relevance to the ground truth. This dataset has been used in visual analytics sensemaking tasks before [45, 98], so we believed it would be a good match for our participants to complete their analysis. In this study, we selected 20 documents from the original text corpus based on the VAST challenge solution. These documents had one main storyline for analysis with two main locations of interest. The documents were about a couple of paragraphs in length on average.

5.3.4 Experimental Tasks

Participants were given a new dataset and task to complete sensemaking on each session. The tasks were assigned in a randomized order and each day the task was loaded into the immersive space as seen in figure 5.1. The three task prompts were as follows:

Sign of the Crescent "From this set of reports, generate a hypothesis about any action(s) terrorists are planning in the near future. The near future means after April 27, 2003, which is the date of the final report you have received. Construct defensible arguments supporting why your hypothesis should be taken more seriously than others that are possible. You must be prepared to issue a warning notice to the new Office of Homeland Security and any other interested offices. At the end of today's session, you will be asked to generate an outline of your findings."

Manpads "From this set of reports, generate a hypothesis about any action(s) terrorists are planning in the near future. The near future means after November 12, 2002. Construct defensible arguments supporting why your hypothesis should be taken more seriously than

others that are possible. Both the DOD and the Department of Homeland Security (DHS) are vitally interested in your reporting immediately on any terrorist threats you believe to be both serious and imminent. At the end of today's session, you will be asked to generate an outline of your findings."

Stegosaurus "Welcome to Alderwood, Washington, a fictitious city in central Washington State. Due to some recent revelations of corruption and conspiracy, the people of Alderwood are left wondering 'What next?' The head of the Alderwood Police Department has a hint towards the answer to that question. An incident from early 2004 (see article IntoxMan.txt) has been nagging him since it happened. Now, with the FBI's attention focused on Alderwood, he has decided to put together a special task force to look into the situation. When he handed you the report he said, 'This incident still doesn't sit right with me. Despite initial reports, no alcohol or drugs were found in the man's blood. That stuff he was spouting sounded like something you'd hear from a cult or something. If not for all the recent goings-on, I might think I was being paranoid. I want this looked into.' Your task is to identify evidence and generate a hypothesis about potentially suspicious activity in the town of Alderwood. At the end of today's session, you will be asked to generate an outline of your findings."

During the outlining stage, participants were encouraged to answer the questions **who, what, when, where, and how** about the threats or suspicious activities; they were also instructed to include any other details they felt were relevant to their hypothesis. Outlines were generated by the participants within an IST note artifact.

5.3.5 Participants

We recruited 14 professional intelligence analysts as participants. The analysts had an average of 12.2 years of professional analysis experience with a standard deviation of 8.23 years (median of 9.5 Years). The age range of our participants was representative of the United States working-class population. We had seven male and seven female participants with the following VR/AR experience levels: Never (8), Once or Twice (4), 3-10 Times (2), and 10+ Times (0). All participants had normal or corrected vision (glasses or contacts). The participants were recruited using word of mouth and email listservs. This study was approved by the Virginia Tech Institutional Review Board.

5.3.6 Measures

We administered a pre-study questionnaire that collected background information on the participants, such as analysis experience and VR experience. During the user study task, we collected both log files and save files. The log files captured all participant actions within our prototype, including participant and keyboard movement once a second and

interaction events (new Note, Search, etc.). The save files captured where all IST artifacts (i.e., documents, notes, labels) were located (x, y, z, yaw, pitch, and roll orientation) once per minute. After each analysis session, we conducted a semi-structured interview with the participants about their analysis strategy and document organization for the session. After the final user study task, we completed a post-study interview to gather insight from the participants on strategies they perceived as most effective, layouts, and suggestions for future features. Finally, we collected a first-person screen recording of the participants' viewpoints while completing their analysis tasks.

5.3.7 Procedure

The user study was broken into three 75-minute sessions. Participants were instructed to schedule their sessions on three separate and preferably consecutive days. Session one began with the formal consent process and the pre-study questionnaire. Next, participants were taught how to use the IST prototype during a ten-minute training phase. Participants were provided a training dataset (distinct from the three datasets used in the main study sessions) and instructed on using all system features. The participants were given five minutes to practice using the features and manipulating the documents within the system.

After the training phase of session 1, and after arrival and quick system review for sessions 2 and 3, participants began their standardized 60-minute analysis task for the day. Participants began the analysis portion of the user study by reading the experimental task prompt (see Section 5.3.4) for the day, and they were informed that they would be expected to produce an outline of their findings using a note within the IST system by the end of the session. Time warnings were given at fifteen minutes, five minutes, and one minute remaining during the analysis portion.

Each session ended with a semi-structured interview while participants remained immersed in the IST system with their analysis documents surrounding them. At the end of session three, participants took part in a post-study interview conducted outside of VR to gain insight into strategies and structures used and collect feedback on the system. Participants were provided an opportunity upon completion of the study to ask any questions of the researchers and provided with contact information if questions or concerns arose.

5.4 Results and Discussion

To address RQ2.1, we looked at the overall spatial structures formed by participants, how these structures differed from session to session and studied the deeper organizational features used for analysis. To address RQ2.2, we examined the videos, interview files, and log-file interactions to understand the strategies used across the sessions. To address RQ2.3, we evaluated the outlines generated by the participants and made correlations between schemas,

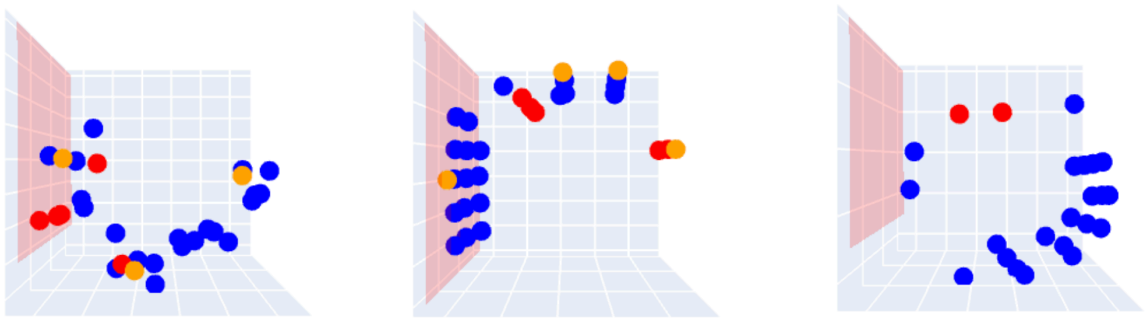


Figure 5.2: From left to right, Semi-Cylindrical, Environmental, and Cylindrical layouts created by P13, P7, and P3, respectively. The red mesh represents the virtual bulletin board, red dots represent notes, yellow dots represent labels, and blue dots represent the documents used during analysis.

strategies, and system-level interactions to understand the sensemaking quality.

5.4.1 RQ2.1: Spatial Layouts and Deeper Organization

Categorizations of Spatial Layouts

Previous studies of the IST approach [59] categorized the types of layouts formed within immersive analytics sensemaking spaces into three types: *Semicirular* ◠, Environmental ⌷, and Planar |. Semicircular has been previously used for layouts in which documents are placed curved around the user. Environmental layouts use the physical boundaries of the space (i.e., floors/bulletin board) for organizing documents in a plane, and planar layouts use planes that exist throughout the space that are not curved around the user. We propose two small changes to the previous layout categorizations in this work. 1) To reflect its three-dimensional nature, we propose that semicircular be renamed semi-cylindrical. 2) We add a distinction between semi-cylindrical and cylindrical ⌭ as supported by recent studies [63, 64, 65, 80], as well by comments from some of the participants. In our new definitions, we define semi-cylindrical layouts as those in which documents are curved around the user but not on all sides, while cylindrical layouts position documents in the full 360-degree space surrounding the user. Our definitions for environmental and planar remain the same. Figure 5.2 shows an example of each type of layout.

At the end of each session, a high-level structure of the space was evaluated using a top-down view such as that seen in Figure 5.1. We found 21 semi-cylindrical layouts ◠, with 50% in session 1, 64.3% in session 2, and 35.7% in session 3. We had eight cylindrical layouts ⌭, with 7.1%, 4.3%, and 35.7% in the three sessions, respectively. We found 13 environmental

	Semi-Cylindrical	Cylindrical	Environmental
Session 1	7	1	6
Session 2	9	2	3
Session 3	5	5	4

Table 5.1: Final layout categorizations for the end of each sensemaking session

layouts , with 42.9%, 21.4%, and 28.6% across the three sessions. In total, 50% of the layouts were semi-cylindrical, 19.05% were cylindrical, and 30.95% were environmental, as seen in Table 5.1. Using a chi-squared test of independence, we did not find a statistically significant difference in spatial layouts used across sessions or datasets.

Spatial Layout Change Over Time

Of our 14 participants, we had eight (57.1%) use at least one different high-level structure across their sensemaking sessions (e.g., P11: $\rightarrow \frown \rightarrow \frown$). This suggests that the participants tried different layouts during their sensemaking as they learned more from using the system for previous analysis tasks. The other six (42.9%) participants used the same high-level structure across all three sensemaking sessions (e.g., P14: $\frown \rightarrow \frown \rightarrow \frown$). Although these participants did not use the space differently across the sessions based on this high-level classification, we did find differences based on deeper organizational features of these spatial layouts, as discussed in Section 5.4.1.

Self-Reported Change Across Sensemaking Sessions

Going beyond the high-level classification of the spatial layouts, during the post-session interviews, participants were asked to comment on their overall spatial structure and if the structures they used during sessions two and three differed from those they used in the previous sessions. The goal was to understand how the participants felt their use of space changed across the sessions. We had ten (71.4%) participants who reported a change in spatial use from session 1 $\rightarrow$ 2 and five (35.7%) participants who reported a change in spatial use from session 2 $\rightarrow$ 3. The changes in perceived space usage can be illustrated by participant quotes:

> *"Much more organized ... Yesterday [session 1] I didn't really have an organization. However, I think it helped me see everything in a way that made sense"* -P4 Session 2 interview

> *"I think on last time I was sort of throwing things wherever to get them out of the way, and then today I took advantage of having the space and using it to organize this time"* -P6 Session 2 interview

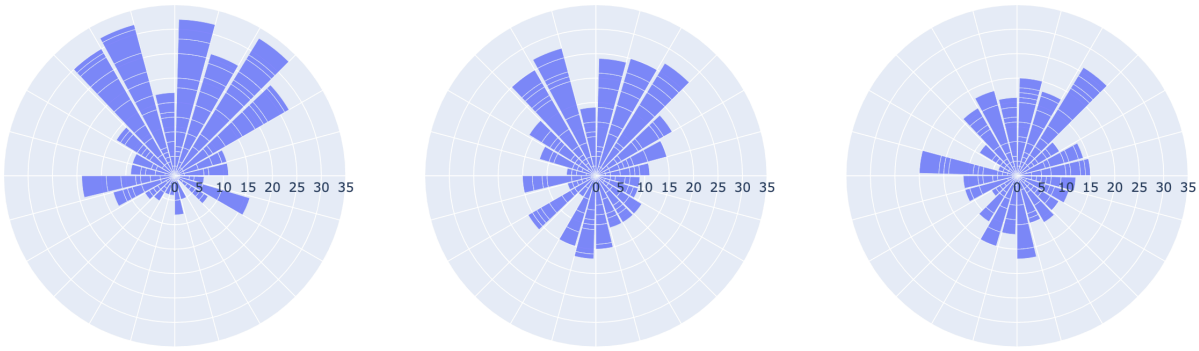


Figure 5.3: From left to right, session 1, session 2, and session 3 angular placement of documents. The direction towards the bulletin board is at the top of the circle. The space is broken into 15-degree bins, and each bar represents the total count of documents across all participants in that 15-degree slice of the space.

"My use of the spatial layout continues to get better as I get more familiar with ... using the tools here." - P14 Session 3 Interview

Angular Space Usage Change over Time

In our high-level classification of the spatial layouts, we saw a shift in more participants using cylindrical layouts across the sessions from 1 → 2 → 3. Additionally, in looking at all layouts, we saw a shift from primarily 180-degree document placement in session 1 to what appears to be more angular space used for document placement in sessions 2 and 3. To take a deeper look at this shift, we analyzed the total angular space used by the participants across the sessions.

To examine the total angular space used for document placement, we: 1) determined the location each participant spent the most time at during a session, 2) translated all document locations to be relative to that location, and 3) calculated the spherical coordinates of each document relative to the center. This allowed us to generate BarPolar charts as seen in Figure 5.3, which represent the total count of documents located in each 15-degree angular bin across the three sensemaking sessions for all participants. The figure shows that in session 1, participants had a strong tendency to place documents in the front 180 degrees of the space. Then as we look at sessions 2 and 3, we see a shift towards using more of the full 360 degrees of the space for document placement. In addition to looking at the angular use of space across sessions, we also looked at the angular use across the three different datasets, but we did not find any distinct differences between datasets for this measure.

This increase in the use of angular space could have occurred for multiple reasons. First, as the participants become more familiar with immersive analysis, they may have been more

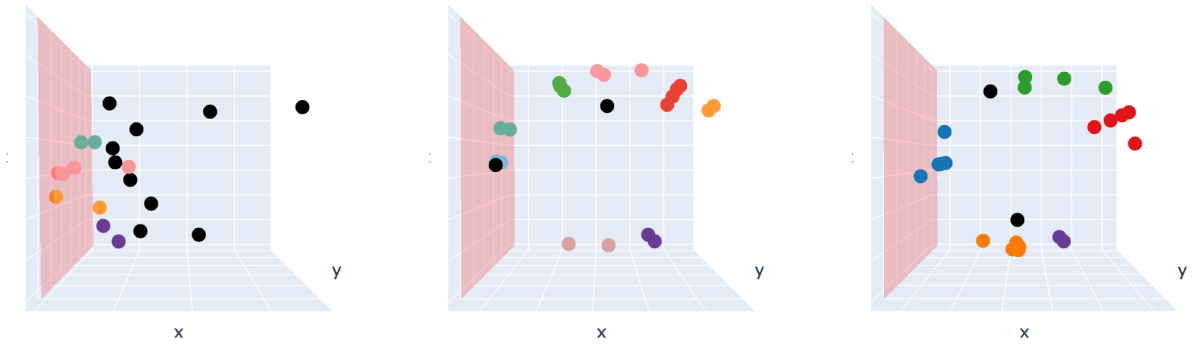


Figure 5.4: From left to right, session 1, session 2, and session 3 clustering using OPTICS for P11's final spatial layout. The red plane represents the bulletin board, the black points indicate documents not assigned to a cluster ("noise"), and the colored points indicate the clusters to which the documents were assigned.

comfortable in trying to use "space to think" during their analysis. Second, based on the interviews, participants came in with new ideas for how they wanted to use the space after their first analysis session, suggesting that they learned from their first session. Another possibility is that participants may have become more comfortable physically navigating and maneuvering within the VR system after an initial session with the system.

Deeper Organization

Clustering Change Over Time

In order to evaluate how participants grouped related documents in space over time, we looked at multiple clustering techniques such as K-Means [67], Hierarchical clustering [69], DBSCAN [30], and OPTICS [3]. Based on the clustering results of running these different techniques, we selected OPTICS (Ordering Points to Identify the Clustering Structure) [3] for use in our clustering analysis since it visually fits the ground truth of the data best. Additionally, OPTICS uses density-based clustering and detects noise (documents that are located too far away from a cluster to be categorized within that cluster). Figure 5.4 shows an example of OPTICS clustering. The goal of this analysis was to identify an overall clustering of each of the final layouts from each session so that clustering evaluation methods such as the Calinski Harabasz (CH) [13] score of the layouts could be used to evaluate the overall organization of the space. The CH score evaluates the goodness of clustering, which is the ratio of the sum of between-cluster variance and within-cluster variance, also known as the variance ratio criterion.

We hypothesized that the CH scores based on the OPTICS clustering would increase over time as the participants became more familiar with analysis in IST. This would indicate well-

defined clusters with a large between-cluster variance and a small within-cluster variance, meaning that documents within the same cluster are close together and far from other document clusters.

We ran OPTICS on each final layout using the following parameters (min_samples=2, metric=l1, and all other default parameters as seen in [72]). We ran a full-factorial analysis of variance (ANOVA) to evaluate the effects of the session, dataset, and the session*dataset interaction on the CH-score from our OPTICS clustering. We found a main effect of the session on the CH score ($F(2,33)$, $p = 0.0114$). Tukey HSD tests for pairwise comparisons found that session 3 had a significantly higher CH score than session 1 ($p = 0.0087$). There was no significant difference between sessions 2 and 3 ($p = 0.4038$) or between sessions 1 and 2 ($p = 0.1399$). The average CH scores were 6.866, 14.882, and 20.186 for sessions 1, 2, and 3, respectively. We found no significant effects of dataset or session*dataset on the CH-Score.

In addition to changes in clustering, there were changes in the use of other organizational strategies over the sensemaking sessions. In previous studies using IST, many participants formed vertical *columns* of documents to encode timelines and form clusters of information. An example of a column can be seen on the left side of the session 3 layout in Figure 5.1. Based on visual inspection in this study, 9/14 participants used columns in their spatial layouts during session one, while all 14 used columns during session 3.

Some participants also organized their notes into a *scratch space*, an area used only for notes that were separate from the space used to organize documents. Based on visual inspection, we found scratch space usage by one participant in session 1, two in session 2, and two in session 3. While there is not a trend with scratch space usage over time, we still highlight it, since this approach to spatial organization has not been reported in previous research related to IST.

The following interview quote highlights how participants were able to use deeper organization within their spatial layouts.

> *"I am actually very proud of this spatial layout. I have it all labeled on top, and I have them all in chronological order but in a way that is easily visible at a quick glance. And I have it broken out by different reporting entity as well."* - P7 Session 2 Interview

Overall, we found that participants utilized the space provided within IST differently as they became more familiar with the tool over multiple sessions. We saw more cylindrical layouts in later sessions, which could be due to comfort levels within the VR system or a willingness to explore different spatial layouts. Most participants also reported that their use of space changed in the later sessions. We also saw more participants using the full 360 degrees around them for document layout as we looked across the different sensemaking sessions and found that the clustering score increased over time. Finally, deeper organizational features were used more frequently during the later analysis sessions. Together, these changes indicate

that participants were learning and adapting to the use of IST as they progressed through the sensemaking sessions, with later sessions tending to exhibit more significant space usage and more organized layouts. We saw the greatest change in the use of space between sessions 1 and 2, with mostly smaller changes between sessions 2 and 3.

5.4.2 RQ2.2: Strategies for Analysis

As approaches to sensemaking are very personal, we expected to see different strategies for each participant in the study. Additionally, we expected to see individuals make changes to their strategy as a result of learning from previous analysis sessions what did not work well and what did. In this analysis, we wanted to look at the participants' strategies from a high-level sensemaking process point-of-view to categorize the overarching strategies during the analysis sessions. Using the interview files and notes from observations of the analysis sessions, we identified five high-level strategies used in our study.

Strategy 1: Read all documents first, then sort the documents into themes. After initial sorting, continue to refine the clusters based on connections between the documents. (Read, Sort, Refine)

Strategy 2: Skim the document set for themes, then begin to sort the documents into groups. After initial sorting, read documents in detail while performing a second sorting. Make refinements to the groupings as needed. "Skim" in this analysis refers to quickly reading over a single document to get a general idea of its contents. (Skim, Sort, Read, Refine)

Strategy 3: Skim documents for themes, then begin searching on recurring themes. Read the documents from the search and, using these results, cluster documents. Use the information from the search to guide the next search. Refine groupings as needed. (Skim, Search, Read, Sort, Refine)

Strategy 4: Sort documents into groups based on a quick scan, then read the documents in detail. While reading, sort the documents into smaller groupings as needed. Refine the space. "Scan" in this analysis refers to a very high-level view of all documents to understand their structure without skimming through them for general details like one does during skimming. (Sort, Read, Sort, Refine)

Strategy 5: Read the documents first, then search for recurring themes. Sort the results from the search and then refine the groupings as needed. (Read, Search, Sort, Refine)

Some of the main differences among these strategies relate to when the "Read & Extract," "Schematize," and "Search for Evidence" stages of the process occur. For instance, in strategy 1, users read the documents first, sorted them in space, made minor refinements to their organization, and then presented their findings in the outline. This strategy is a very linear sensemaking process compared to strategy 2, where participants started by skimming the

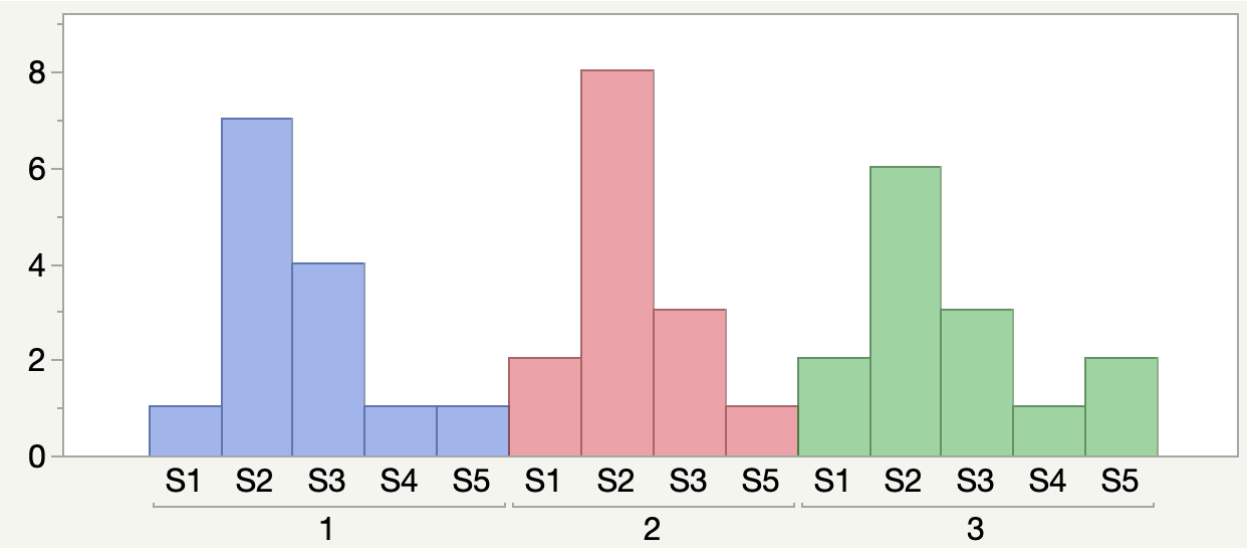


Figure 5.5: Total count of strategy usage across the sessions. Blue is session 1, red is session 2, and green is session 3.

documents, then sorting the documents into rough groupings. Then the documents were read in detail and sorted again through a refinement stage while the participants prepared their hypotheses. This approach involved more iterations up and down the sensemaking loop. Strategy 3 and Strategy 4 were also more iterative high-level strategies, while Strategy 5 was more linear. Our study's strategies are similar to those defined by Kang et al. [48], with our S1 mapping well to "Build from Detail", S2 mapping to "Overview, Filter, and Details", S3 mapping to "Hit the Keywords", S4 mapping to "Overview, Filter, and Details" with an initial triage stage at the beginning, and lastly, S5 mapping to "Find the clues, follow the trail."

Figure 5.5 shows the counts of strategy uses per session in our study, with Strategy 2 being the most popular strategy in all sessions. Using a Chi-Squared test of independence, we found no statistically significant difference in using different strategies across the sessions or the different datasets. In order to evaluate interaction-level differences among the strategies, we ran Student's T-Tests on system interaction data collected in the log files. We found that participants using S2 walked farther than those in S1 ($p = 0.0259$), with an average of 151.28 meters compared to 101.559 meters of total movement. We found that S3 ($p = 0.0278$) and S5 ($p = 0.0311$) users executed more searches than those who used S1. Lastly, we found that S1 users moved documents more often than S2 ($p = .0029$), S3 ($p = 0.103$), and S5 ($p = 0.0463$). One interpretation of these findings is that the participants who used S1 spent more time reading and moving documents around than using features to help them make connections, such as the search feature or physically navigating the space to help them understand the overall dataset. Another interpretation of these findings is that since S3 and S5 executed more searches than S1, they better understood where the documents connected

on a thematic level and, therefore, needed less time moving documents to create meaningful organization. Additionally, as S2 navigated more during their analysis than S1, this could suggest that S2 used physical navigation via walking around the space to aid in referencing documents during their analysis instead of using the controller to bring documents to them for examination.

During the post-session interviews, participants were asked to comment on their overall strategy and if their strategies during sessions 2 and 3 differed from the ones they had used in the previous sessions. Ten participants reported a change in strategy from session 1 $\rightarrow$ 2, and six reported a change in strategy from session 2 $\rightarrow$ 3. We found that the majority of participants self-reported that their strategies changed after session 1. These self-reported changes capture high-level perceived strategy differences (e.g., reading all documents first $\rightarrow$ sorting documents first) and system-level interaction changes in strategy (e.g., notes, labels, and search usage). Overall, participants utilized different high-level strategies for analysis as they became more familiar with using IST, as we expected. While these changes in strategy may also show a learning effect, we believe that these changes provide insights into how IA systems, such as IST, would be used in a real-world setting by providing the participants time to learn how the system could truly benefit them during their analysis process. A few quotes highlighting strategy changes from the participant interviews can be found below.

"I don't think I really had a strategy the first time ... I kept jumping around on which piece of information I wanted to group by, and this time it seemed to make more sense to make more of a timeline strategy." - P6 Session 2 interview

"Yesterday, I read every report before I did the buckets, and I kind of saw that that was a waste. So today I skimmed them more to kind of get the ideas together ... I felt more comfortable with the controls and I don't think I reread as much today." - P8 Session 2 Interview

5.4.3 RQ2.3: Correlations with Sensemaking Quality

To this point, we have presented spatial layouts and high-level strategies used during analysis. However, we also aimed to understand how layouts, strategies, or system-level interactions correlate with the quality of the sensemaking task.

In order to evaluate sensemaking quality, we scored the outlines generated by participants based on the ground truth of the dataset. The rubrics awarded points for every correct piece of information presented and specifically focused on the **Who, What, When, Where, and How** questions the participants were guided to include in their outlines. Where a point was awarded for every correct piece of information that was reported on. The primary experimenter scored the outlines in a randomized order across all three datasets. Each

	Average	Standard Deviation
S1	0.2250	0.0212
S2	0.6392	0.1537
S3	0.5261	0.2445
S4	0.7400	0.0000
S5	0.3483	0.3483

Table 5.2: Average score and standard deviation for each strategy

rubric had a variable number of points that could be received (Crescent - 25, Manpads - 26, Stegosaurus - 31), so we report normalized scores where 1.0 represents a perfect score.

We ran a full-factorial ANOVA to evaluate the effects of Session, Dataset, and the Session*Dataset interaction on the outline scores. We found no significant effect of Session or Session*Dataset; however, we did find a main effect of the dataset on the score ($F(2,33)$, $p = 0.0004$). Using an LSMeans Differences Tukey's HSD Test, we found that Sign of the Crescent ($p = 0.0020$) and Manpads ($p = 0.0007$) scored higher than Stegosaurus. Due to the apparent difficulty of the Stegosaurus dataset, we removed it from the following correlation analyses.

We also ran ANOVAs to evaluate the effects of Session and Dataset on all other variables listed below (spatial layouts, strategies, and interactions). We did not find any effects of Session, Dataset, or Session*Dataset on any of these variables, and therefore suggest that the correlations presented below are not influenced by dataset or familiarity. Thus, we can have some confidence in a general relationship between sensemaking quality and specific layouts, strategies, and interaction patterns.

Spatial Layouts and Quality

The mean score for cylindrical was 0.5535 (sd = 0.2162), environmental was 0.5689 (sd = 0.1320), and semi-cylindrical was 0.5394 (sd = 0.2646). We used a Student's T-Test (alpha = 0.05) and found no significant effect of the layouts on scores.

Strategies and Quality

The mean score and standard deviation for each strategy can be seen in table 5.2; since we were interested in evaluating the difference between iterative sensemaking and linear sensemaking, we combined the strategies into two categories to see if these different approaches affected overall score on the sensemaking task. Using a t-test, we found that the iterative sensemaking strategies (S2, S3, and S4) scored higher on the sensemaking tasks than the more linear approach to sensemaking (S1 and S5) ($p = 0.0094$).

Interactions and Quality

We looked at all log-file interactions (searches, grabs, labels, notes, highlights, text entered, and copy/paste) as well as calculated features (keyboard and participant movement, time spent interacting with documents, and distances between documents) in an attempt to understand how these correlate to the quality of the sensemaking.

We found two interesting correlations that we want to highlight here. We found a positive correlation (0.4596) between the score and the number of **new notes** created during analysis (p=0.0139). We wanted to understand the **"spread"** of documents across the provided space (i.e., the amount of space used to lay out the documents). To calculate the spread, we looked at the total distance between all documents in the space. Then we normalized by the total number of artifacts (docs, labels, and notes) within the participant's layout. We found a positive correlation (0.4703) between this factor (Sum Distances / Number of Artifacts) and the score (p=0.0116), showing that a layout with greater spread correlates with higher-quality sensemaking.

5.4.4 IST Analysis Best Practices

Based on the analysis conducted in this study and the findings presented above, we propose a set of best practices for effective analysis within IST-like systems. Our study suggests that these guidelines can help users perform effective sensemaking in IA systems.

Use a spatial organization schema that works best for the task and analyst preferences There are many ways to encode meaning in a system like IST. Having the availability to organize in 3D allows for new opportunities to organize sensemaking schemas. We saw no differences between spatial layouts in terms of the quality of sensemaking. Thus, we suggest that the overall spatial layout should match the document set and the analyst's preferred style.

Use a more iterative approach to sensemaking We found that participants with more iterative approaches to sensemaking scored higher than the more linear approaches. Based on this and user feedback in the final interview session, we believe frequent iterating through the sensemaking stages leads to better analysis both in and out of IST.

Use more space for document layout As participants became more familiar with the tool, they became more organized over time. Additionally, the participants who used more of the provided space (better spread between documents) scored better on these sensemaking tasks. We suggest that using more space to create meaningful organization benefits the overall sensemaking quality. In future iterations of IA systems, additional organization features such as semi-automated clustering [88] or semi-automatic column detection could provide scaffolding for users to create meaningful structures.

5.5 Limitations

One limitation of this work is the text-entry system. For this study, we used a tracked keyboard and table that participants could wheel around the space with them. While we removed some of the foam in the headset to assist the user in seeing their fingers on the keyboard, some participants still struggled with text entry, which could have led to less detailed outlines or fewer notes. In future work, a system like the portal proposed by Giovannelli et al. could be used [35]. Moreover, as a result of using the VIVE Pro headset, the text-readability when documents were placed across the 3x3 meter tracked space decreased due to some blurriness. While we don't believe this influenced our results, we suggest that future work use higher resolution displays to enhance text-readability from across the tracked space.

Another limitation of this work is that the Stegosaurus dataset differed from both Manpads and Crescent in document length and style, which could explain the difference in scores we observed, thus limiting our ability to evaluate the effects of session and other factors on scores.

Finally, a small participant pool limited our statistical analyses, resulting in less power. We present all statistically significant findings with caution and provide directions for future research that should be verified.

5.6 Conclusions

This work explored how learning and familiarity influence sensemaking in immersive analytics. We explored three main research questions that focused on spatial structures, strategies, and correlations for performance. We found high and low-level spatial structure and strategy changes across the different sensemaking sessions. Together, these indicated that as the participants became more familiar with the tool, the participants adapted and learned how to analyze within our prototype IST. Lastly, our correlation analysis found that iterative sensemaking, note usage, and document spread were key indicators for effective sensemaking. Using what we learned in our exploratory analysis, we suggest guidelines for sensemaking in IA systems.

We also covered many types of analysis within this paper to explore users' sensemaking in IA systems. The analyses included were not limited to spatial structure categorization, deeper organization features, clustering analysis, degrees used, high-level, and strategy categorization. We believe these analyses provided us with many insights into our participant's sensemaking and highlight these novel analyses to be used in future work for IA sensemaking. One of the future directions of this research will be implementing new organization features intending to understand how these features affect the sensemaking task. We will specifically look at new features that would promote some of the best practices described

above. To promote spatial organization schemas, we believe some automated layout features can be developed to assist the users in the overall organization. Additionally, additional organizational features can be developed to encourage more space usage to help users offload more sensemaking into the environment. Some of these features include the clustering of documents, which allows moving sets of documents together as a group. Another idea would be to provide assistive linking features to help analysts tie information within and across clusters, allowing for the ability to have cross-space references when needed.

Chapter 6

Designing for Best Practices in Immersive Space to Think

IST has shown promise in supporting intelligence analysis and sensemaking by providing a vast space to think, allowing users to organize documents in new and meaningful ways. However, our prototype thus far has had limited features for assisting in the sensemaking task. In our prior work, we have provided only basic analysis tools and this vast space for organization. By giving no additional organization features, we have forced our participants to focus part of their cognitive load on organizing their space and creating meaningful layouts during their analysis. If we provide novel organization features for our participants to form meaningful schemas, we could help offload some of the overhead of the analyst during their sensemaking process into the system through these additional organizational features.

In chapter 5, we gained an understanding of the initial strategies and spatial organizations used with IST and if they persist after participants become more familiar with immersive sensemaking analysis. Overall, we found that users refined their strategies when given the opportunity and became more organized in their use of space. Additionally, we evaluated the overall strategies, spatial organizations, interviews, and sensemaking quality via the outline score to present a set of **Best Practices** for using IST and how we can design features within immersive analytics systems to help support these best practices. Those findings can be summarized in table 6.1.

In this chapter, we aim to address RQ3 - How can we support best practices for immersive sensemaking with additional organizational features? To answer this question, we developed a new set of organizational features that are designed to support users during the schema-making stage of the sensemaking process, as seen in figure 1.1. Then, to understand how these features affect the sensemaking process, we conducted a user study to evaluate the overall effect of these new features.

Best Practice	Designing For Best Practices
Use a spatial organization schema that works best for the task and analyst preferences	Automated layout features to assist with high-level organization
Use a more iterative approach to sensemaking	Develop assistive features that encourage participants to begin externalizing their cognition early into the sensemaking process
Use more space for document layout	Develop a new set of organizational features to encourage increased space usage such as clustering, assistive linking features, suggested alignment features, etc.

Table 6.1: Best Practices and How to Design to Support Them

6.1 Feature Design[1]

We developed the new features for IST in two stages. In the first stage, we completed the initial prototyping of the new features. Then, using these initial designs, we ran an expert feedback session with professional analysts to gather feedback on the new features and gathered feedback on how we could improve these features to better support the analyst's sensemaking needs. After running the expert feedback sessions, we then enhanced these new features based on the feedback gathered. In this section, we present the initial prototyping, expert feedback sessions, and a summary of the feedback we received.

6.1.1 Initial Prototyping

Clustering

As users complete sensemaking within IST, they naturally form groups of information based on document relationships. For example, columns/timelines are common ways users organize information during analysis. Building on this, we wanted to provide a way for users to explicitly group related items while providing a visual artifact to encapsulate that information, which could be achieved through explicit clusters. By providing clustering capability, users can make deliberate decisions about document placement within the analysis space, making the relationship between documents more clear. Furthermore, clustering can enable users to move entire groups of information together rather than relocating each document individually, thereby streamlining the organization and analysis process.

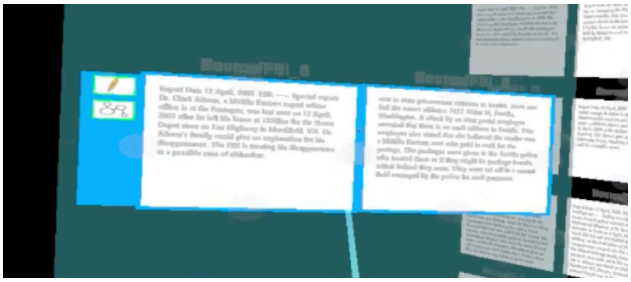
Prior work in IST examined how clustering could be implemented into the system during an analysis task [88]. The users completed three image classification tasks in this experiment: Animals, Types of Transportation, and Foods. This study evaluated three different techniques for interaction - freestyle (no clusters), which represented the baseline, overlap,

[1]All figures in this section are presented as screen-shots with black backgrounds. While these images have black backgrounds, the prototype used the AR Video Pass-Through on the Varjo XR-3. For privacy reasons, the screen recordings were captured in the Unity Editor Screen, which does not include the AR-Video Pass-Through. This also leads to low-quality video capture, resulting in decreased readability.

and proximity. This study found that while the semi-automated clustering or proximity technique was faster and preferred for simplicity and ease of use, the manual (overlap) cluster technique allowed the users greater control. Due to the complexity of sensemaking with text-based data, we believe there is more nuance to the location of a document than may have been seen in the previous work by Tahmid et al. [88]. In this prior work of clustering pictures of animals, methods of transportation, or food, it is easier to think of explicit categories into which those items can be clustered. However, with text-based sensemaking, such as intelligence analysis, a document may relate to more than one cluster, meaning it belongs between clusters. Placing a document between clusters would be time-consuming in a proximity-based semi-automated clustering mechanism. For this reason, we choose to utilize the overlap clustering method.



(a) Overlap of documents turns them yellow to indicate that upon release, the documents will merge to create a cluster.



(b) Outcome of releasing a document from part A. A cluster container has been created with the two documents moved, so they do not occlude the content of the documents.

Figure 6.1: Cluster Creation Overlap Technique

Creation In overlap, a user can create an explicit cluster by overlapping two documents. If the documents are touching, their background turns yellow to indicate that a cluster will be made if the document is released. Upon releasing a document with a yellow background, the documents are placed into a semi-transparent cluster container in blue oriented towards the user. Figure 6.1 shows an example of the clustering interaction.

Insert into a Cluster Once a cluster is created, users who want to add a document to an existing cluster also use the overlap technique. The user will pick up a document and overlap

it with the cluster container, in which case each background will turn yellow to provide the user with feedback. Then, upon release, the document will be added to the cluster.

Merge Clusters Merging clusters occurs using the overlapping technique as well. Once a user has two clusters they want to merge, they must overlap them. Then, the two clusters will turn yellow to indicate to the user that the cluster merge will occur if released. Upon release, the documents within the two clusters are merged into one larger cluster.

Cluster Deletion To remove a cluster from the analysis space, a cluster must be down to two remaining documents. Then, upon removing the second to last document from a cluster, the system will delete the cluster from around the remaining document. This prevents clusters of one to exist in the environment.

Alignment

Prior work with clustering in IST, [88], developed a feature to prevent occlusion on cluster creation, insertion, and merge. This means that the user places two documents on top of one another within the cluster, the cluster container will grow and move the two documents so they are no longer overlapping (i.e., stacked → side-by-side) as seen in figure 6.1. This feature allows the users to quickly add information into a cluster without worrying that the added information will occlude any other information within the cluster. Additionally, the users can refine the placement of the documents within the cluster by moving them around in the container.

However, these features do not provide the user with quick organization techniques, such as specialized alignment features to assist in tidying up one's space or organizing based on the content of the documents. However, one challenge in providing a feature to help the users in tidying up the space is determining when to provide alignment and which documents to align. For this implementation, we decided to focus on a *manual* approach to alignment, and we wanted to prevent documents from being rearranged without the user explicitly telling the system to rearrange the documents. To implement this *manual* approach, we created a series of buttons that the user must select before any documents are rearranged. Next, we needed to decide which documents should be rearranged and aligned for the users. One approach to determining which documents to align would be to create a 3D lasso interaction for the user to select a set of documents to align. However, since we already included a clustering feature in our design to provide additional organizational features, we decided to leverage these clusters as the "selection" that the analyst would want to tidy up with the built-in alignment feature.

To explain how the cluster alignment features seen in figure 6.2 work, we use an example with a mock cluster and documents. The starting state of the alignment features listed below is seen in figure 6.3.

Row In this design of the auto-alignment feature, the user used the ray from the controller



Figure 6.2: An IST view of the cluster alignment feature with the Row alignment applied. The menu for selecting a quick alignment can be seen in the upper left-hand corner of the cluster. The options from top to bottom are Row, Column, and Timeline.



Figure 6.3: An example cluster starting point for showing the different alignment options.



Figure 6.4: Row Alignment Cluster Example

to hover over the button on the cluster menu. Upon hovering over the button, the button is outlined in a light blue color to indicate to the user that this is a clickable feature. Then, upon selecting the row button, the documents are organized into a row. The boundaries on the row are four documents across before a new row is added below that row, and so on. The order of the documents in the row is decided upon by looking at the (X, Y) coordinate of the documents within the 2D panel (cluster). To determine how many rows, first, the number of documents within the cluster is calculated. Then, we can get the total number of rows as $RowCount = \frac{NumberOfDocumentsInCluster}{4}$. We sort the documents based on their Y coordinates to determine which ones will be in which row, where the four highest documents are in the first row, then the next highest, and so on, then the X coordinate is used to determine the location of the documents from left to right in the row, from lowest x to highest x — an example of the row alignment figure can be seen in figure 6.4.

Column After selecting the column alignment button, the documents are organized into a column layout. The boundaries on the column are four documents high before a new column is added to the right of that column, and so on. The order of the documents in the column is decided upon by looking at the (X, Y) coordinate of the document within the 2D panel (cluster). First, the number of documents within the cluster is calculated to determine how many columns. Then, we can get the total number of rows as $ColumnCount = \frac{NumberOfDocumentsInCluster}{4}$. We sort the documents based on their X coordinates to determine which ones will be in which column, where the lowest x-coordinate documents are in the first column, then the next lowest, and so on. Then, the Y coordinate is used to determine the location of the documents from top to bottom in the column, from the highest y at the top of the column to the lowest y at the bottom. An example of a column alignment figure can be seen in figure 6.5a.

Timeline After selecting the timeline alignment button, the documents are organized into a timeline layout. The boundaries on the timeline layout are four documents high within a column before a new column is added to the right of that column, and so on. To decide on



(a) Column Layout Applied (b) Timeline Layout Applied

Figure 6.5: Column Alignment Cluster Example

the order of the documents, the text content is parsed for the report date. For example, in FBI_15, the report date is 15 April 2003. The documents are sorted from the earlier report date to the most recent and then placed in the columns where the earlier report date is in the highest location in the leftmost column, the next earliest below that document, and so on. If there is a tie with the document's report date, the document with the alphabetically earliest title breaks the tie. Figure 6.5b shows an example of the timeline layout.

Linking

The last set of features we designed for the expert feedback session were related to *Linking*. During previous post-session interviews with professional intelligence analysts, interest in being able to create links within the analysis space was voiced. Some quotes from the analysts can be seen below:

"Having the ability to go beyond clustering and being able to connect with string different elements of information...Visually connecting information outside of just clustering items together could be helpful" - *P11, Exploring Professional Analyst Strategies in Immersive Space to Think*

"I wish I had strings to attach things to each other...even in the 3D space following the string across the room"- *P7, Uncovering Best Practices in Immersive Space to Think*

"I was thinking it would be nice to be able to draw some sort of either network map or the labels or with the small box. It was hard to tell who was connected to whom, so having that sort of linkage that you could visually see as you are using the space would be really helpful, especially as you are trying to outline your report." - *P9, Uncovering Best Practices in Immersive Space to Think.*

To fill this gap, we developed a set of linking features to gather feedback on during the expert feedback session. The first iteration of these features is described below.

Manual Linking In our design of the linking feature, we created four "anchor" points on each IST artifact (document, label, and note). The anchor points are represented with a mostly transparent sphere on the top, right, bottom, and left-hand side of each of the items, as seen in figure 6.6. To create a link between two artifacts, the user casts the ray from the controller over the top of one of the spheres, turning the sphere temporarily opaque to provide feedback to the user. Then, the user can select the trigger on the tracked controller, and a link will spawn from that sphere to the controller. This link is now in an "editing" mode. Next, the user can move the controller over the "anchor" point of the artifact they want to link to. Then, using the trigger on the tracked controller on another "anchor", the link will jump from the controller (where the temporary endpoint is placed) to the document. There will now be a link that is created between these two documents, as seen in figure 6.6.

To help keep the links tidy and prevent occlusion, we wanted to provide a way for the links to update based on where the two documents were located relative to each other. For instance, without providing some tidying, if the analyst wants to move the document CIA_4 in figure 6.6 above the other document, the existing link would draw over the document to connect the two. To prevent this, every time an artifact within IST is moved, we run a background process to check if it has any links, and if it does, to find the shortest path between all anchor points on the current document (the one being moved) and the documents that it is linked to. This allows us always to have the shortest lines between the two artifacts, helping to keep the links tidy.

Discovery Linking Thus far, all of the features we have described in this section have focused on providing additional support in the *schema* stage of the sensemaking process. We believe these new features will provide the users with new ways to organize the information and will assist them in offloading cognition into the immersive environment. However, to provide additional support through the sensemaking process, from foraging up to presenting findings, we wanted to develop an additional feature with links to provide support in the

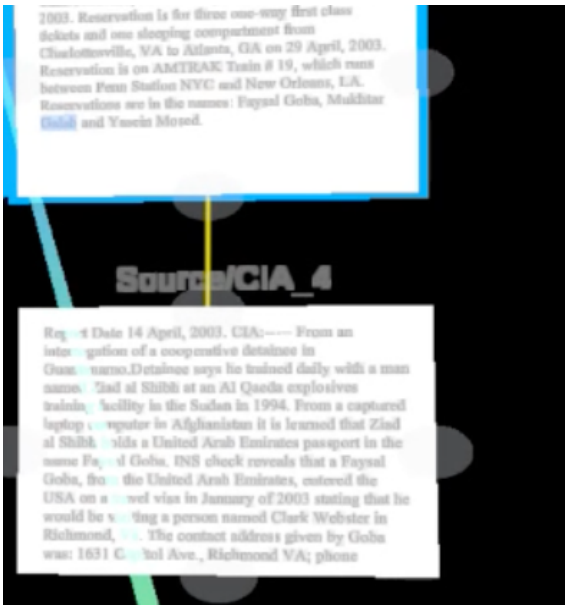


Figure 6.6: Example of a Manual Link Created During Analysis.

foraging loop. In a previous user study interview, we received feedback from a professional analyst who stated "I wish I could have all the data look at itself and pull any similar information and highlight it for me. Like, look for any term found in all the other documents." - *P8, Investigating Professional Intelligence Analyst Sensemaking in Immersive Space to Think*. Building on this idea of being able to pull any similar information with highlighting along with the linking feature, we developed a feature we call **Discovery Linking**.

Opening A Discovery Link To create a discovery link, a user can select text from a document using the traditional IST text selection method of hovering over the text with the ray cast while holding an input button on the controller. After selecting text from the document, the user can select the button on the user interface panel attached to the document to create the discovery link. Upon selection, a discovery link will open surrounding the document where the button was pressed. Figure 6.7 shows an example of an open discovery link.

Upon selecting the discovery link button, the IST system searches the document set for any words/phrases matching the selected text in the spawning document. Then, any documents containing that matching word will be "pulled" from their current position and placed around the spawning document in a series of concentric circles while ensuring no overlapping documents exist.

Persistence of A Discovery Link / Manual Linking from Discovery Links For this iteration of the discovery links, we imagined that this feature would assist the user in finding other information relevant to their analysis and bring it close to other documents with that related term. However, we did not imagine that this feature would be used to move

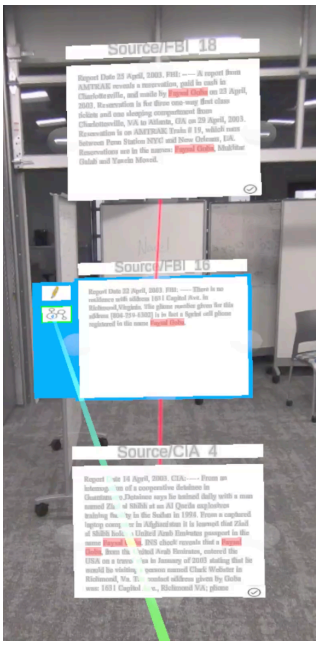


Figure 6.7: Example of a Discovery Link Executed on the name "Faysal Goba".

documents in the analysis space permanently. For this reason, the discovery link feature is currently implemented as a toggle, where the discovery link is opened on the first press and then closed on the second press of the button. Upon closing, all documents brought forward during the discovery are sent back to their last known location in the analysis process. However, suppose the user finds a document they wish to keep nearby after undoing the discovery link. In that case, the analyst can turn this link into a yellow or permanent link using the check box in the bottom right-hand corner of the document, as seen in figure 6.7. Upon selecting the checkbox, the red link will turn yellow. Then, upon closing the discovery link, this document will remain in its new location.

Discovery Links Interaction with Clusters When designing this feature, we needed to consider how the Discovery Links should interact with the documents that have been clustered. We had two options to choose between when designing the feature. First, we could pull a document from a cluster and place it around the spawning document, as with all other documents. Then, upon closing the discovery links, the document would be placed back in the cluster as before. However, when we thought more about clustered documents, we felt like pulling a document from where the analyst had already "sorted" it into a schema broke the idea of what a cluster represents. Therefore, instead of moving documents out of the clusters, we opted to leave the document in the current cluster location and still draw a line connecting the spawning document and the clustered document. This ensured that a document wasn't moved from its intentional location and that the user could still see the connections from the current document to the clustered one.

6.1.2 Expert Feedback Sessions

The features we detailed above aimed to be the first iteration of the design of additional ways for the users to organize the information in the immersive environment. In this design stage, we decided how we believed the features should interact based on our previous interviews with IST participants and our understanding of how users complete sensemaking within IST. Then, we conducted an Expert Feedback Session with Professional Intelligence Analysts to gather feedback on our design implementations and brainstorm additional ideas.

System Features

We wanted to focus the feedback on these newly implemented features described above. For this reason, we elected to turn off some of the other IST features to shorten the tutorial/on-boarding time. With this in mind, we ended up with an IST prototype with the following features: Note taking, label making, clustering, clustering alignment features, linking, discovery linking, searching, and highlighting.

Feedback Task

In order to get the analysts familiar with all of the features within the system, the participants were given a fictional intelligence analysis training dataset to complete some analysis on. This dataset was the *Sign of The Crescent Dataset*, described in greater detail in section 3.3.4. The Sign of the Crescent Prompt document was still loaded into the system to orient the user to the task they could complete with the given dataset. However, the participants were made aware that the session's goal was to provide system feedback and not to conduct an analysis of the given dataset. We provided the users with 25 minutes to complete some initial analysis within the system and attempt to utilize all of the different features provided to them. Time warnings were given at 12, 5, and 1 minute remaining and reminders to use all the features provided.

Feedback Participants

We recruited nine professional intelligence analysts as participants in this feedback session. The analysts had an average of 10.3 years of professional analysis experience with a standard deviation of 8.7 years (median of 10 years). The age range of these participants was representative of the United States working-class population. We had three female and six male participants with the following VR/AR experience Levels: Never (1), Once or Twice (3), 3-10 Times (2), and 10+ times (3). All participants had normal or corrected vision (glasses or contacts). The participants were recruited using word of mouth and email listservs. The Virginia Tech Institutional Review Board approved this expert feedback session.

Semi-Structured Interviews

The semi-structured interview was the focus of the expert feedback session, and we expected it to take around 30 minutes to complete. The interview questions focused on specific questions about linking and clustering features. The semi-structured interview ended with an I like.../ I wish.../ What if... exercise, where the user was asked to phrase feedback on the features and design using these sentence stems [71]. This exercise encourages users to convey things they liked about the system with positive feedback, share ideas on what they wish to improve, and make suggestions about the system that may not be directly related to the prototype. The semi-structured interview questions are detailed in appendix C.

6.1.3 Results from Feedback Sessions

To gather the comments from the feedback session, all of the interview files were transcribed for analysis. Then, these transcripts were broken down into comments relating to the specific features they were related to (e.g., clustering alignment, linking, etc.). We then further broke down the comments into positive feedback and suggestions. The goal here was to develop an understanding of what worked well for the analysts about these new features and how we could modify them specifically to better support their needs. To gain specific insights into the most wanted modifications to our new features, we grouped them by similar suggestions and counted the number of times certain suggestions to our current implementation were made. We have broken down the feedback we received around specific features below:

Clustering

The overall feedback we received on the clustering features was very positive. The participants like the fact that they could place "a bunch of related documents in the same cluster" - P1. Additionally, P7 highlighted that "It also helped me with making a conscious analytic decision about what I was doing there and then be able to walk away from it and come back later and understand why/what decision was made earlier." Overall, we received 4 main suggestions surrounding the clustering. First, the analysts wanted the ability to change the color of the clusters (5 participants); for example, P4 stated, "If I were to have a bunch of different colors, I could use the colors to denote lots of different information like confidence." Second, the participants wanted a way to label the clusters created (4 participants). Thirdly, the participants expressed interest in a way to minimize or resize clusters to fit custom dimensions (5 participants), with P7 stating "I think being able to minimize the documents down to their labels would be really important". Lastly, a feature that the analysts also expressed an interest in was the ability to search for keywords/phrases within a cluster (4 participants).

Automated Layouts

The feedback received on the automated layouts was very positive. Overall, the timeline layout option was the most helpful to the experts. Additionally, the analysts liked the fact that they could place a bunch of documents relating to a concept in the cluster and clean the cluster up afterward. For example, P4 stated, "I had added a bunch of information into the cluster, so I used these features to basically re-organize the mess and make sure that it was orderly." We received two main comments on the automated layouts within the cluster. First, the participants wanted more layout options (e.g., grid layouts or hierarchical layouts (3 participants). Second, P1 raised a point about adding a feature to automatically sort a document when added into a cluster with an automated layout, stating that "I would definitely turn it on as I was adding to the cluster, then turn it off to create more nuanced organizational structures within the cluster."

Linking

Manual Links

The overall thoughts on the linking features were positive, with P8 stating, "I think that the linking feature is the most useful one...I would use this in a real environment. I think the linking feature is really good for showing many-to-one relationships." Additionally, P6 highlighted that "I think my inclination was to use the yellow linking (manual linking) as sort of like a temporary thing. I think if I were doing a more thorough analysis, I think the yellow link helped me draw connections while finding them, and then the cluster was more of a … I have decided which relationship matters, and now I will cluster them." We received two main suggestions regarding the manual linking features. First, as expected, all the participants wanted the ability to edit links using features such as color, deleting, or temporarily hiding them. For instance, P4 stated that they might use color to denote specific connections like "green, meaning address, red, meaning phone number, etc." or utilize the color to denote overall confidence in the links created "like a green link means high confidence, red mean explore more." Second, some analysts requested the ability to link information to clusters or link clusters to other clusters (6 participants).

Discovery Linking

We did not receive many suggestions on improving the discovery linking features. Overall, the analysts felt the feature helped with efficiency, with P4 stating really liked the link search (discovery linking) feature. I thought it made things super efficient and easy." Additionally, P6 stated, "Discovery linking feature I think that is really powerful...I think the reason I like it so much is it keeps you in the same context and workflow." P4 commented that adding metadata within discovery links could help make the feature more robust and that having the ability to execute follow-on discovery links could be beneficial in the analysis process as well.

The goal of collecting this feedback from the analysts was to understand 1) how these features currently benefited their analysis process, 2) how we could modify these features to make them more helpful for analysis, and 3) home in on where to make modifications for a future user study aimed to understand the effects of these new organizational features on the sensemaking process. Before diving into the modifications made to the new organizational features, we first present the experimental task for the user study.

6.2 User Study Conditions and Experimental Task

This user study aimed to understand how the new organizational features affect the overall sensemaking within IST. To measure these effects on the sensemaking process, we wanted to evaluate how each new feature influenced the overall sensemaking process. To achieve this and be able to make these comparisons, our user study was a within-subjects design with four conditions. The 4 conditions were: **Freestyle IST (Baseline)** - IST with no new features, **IST Clustering** - IST new clustering related features, **IST Linking** - IST plus the linking features (both manual and discovery), and **IST Clustering & Linking** - IST with the new clustering & linking features.

In the previous user studies presented in chapters 3, 4, and 5, we have focused on having our users complete fictional intelligence analysis sensemaking tasks, which vary in size from 20-40 text-based documents. However, some challenges with these tasks are finding 1) similar size document sets, 2) similar difficulty tasks, and 3) matching task varieties. For example, in chapter 5, we had our analysts complete three different tasks: Sign of the Crescent (shortened), Manpads (shortened), and Stegosaurus (shortened). While the document sets were curated to have a matching number of documents, the tasks differed slightly across these datasets. For instance, Sign of the Crescent and Manpads are fictional intelligence analysis training tasks. Stegosaurus was a previous VAST challenge dataset that was detective-based but fundamentally different from the other two tasks. For example, the style of the documents was different from 1-2 paragraphs in length on average to news article length on average. In comparing task performance across the three datasets, we also found that the Stegosaurus dataset scored statistically significantly lower than Sign of the Crescent and Manpads, suggesting that these tasks are not equal in difficulty. In addition to these tasks being challenging to match difficulty levels, these tasks take a long time to complete. For example, in the previous studies, the full Sign of the Crescent was designed to be completed over the course of 3 different sessions, taking about 3.5 hours to complete. Then, in the modified smaller tasks, we gave the analysts 60 minutes to complete the task.

Therefore, to keep the same task structure for this study as our previous work, we would have needed to curate two additional text-based datasets similar to Sign of the Crescent and Manpads in difficulty and in time for task completion. Then, we would also need to schedule the participants for four separate 1.5-hour user study sessions for each user to experience all of the experimental conditions. While this could be doable, this becomes increasingly

challenging. Additionally, if a user chooses to drop out of the experiment after multiple sessions, we would lose a lot of data. For this reason, we selected a different analysis task in this comparative study.

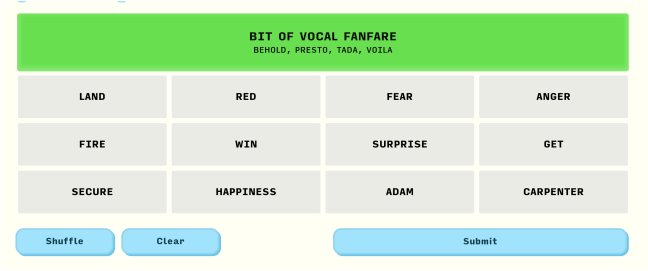
For this study, we wanted to select a text-based sensemaking task that could be completed in 15 - 25 minutes to allow the participants to schedule a 2 - 2.5 hour session to complete the entire user study. While we could attempt to curate the existing Sign of the Crescent / Manpads datasets to contain a smaller number of documents and break them out into four separate tasks, we may lose some value of task difficulty and vastly decrease the number of artifacts used for analysis (documents). For this study, we modified a sensemaking game called Connections [2]. Connections is a game where the goal is to organize a set of 16 words into four groups based on the underlying *connections* of the words. An example starting state of the connections game can be seen in figure 6.8a. In this game version, users can select four cards from the board that they think are connected and submit them to the system. The user has four wrong attempts before they lose the game. In this version, after each correct answer is submitted, the system provides the user with the title of the connection, for example, *Bit of Vocal Fanfare*, as seen after submitting the correct connection in figure 6.8b. An example of a completed Connections game board can be seen in figure 6.8c. In this version of the game, where the user submits their answers one connection at a time, the colors represent how challenging the connections were to find relative to the others in this board, with easiest to find to hardest to find in the order yellow, green, blue, then purple. In this game version, after submitting a correct answer, the number of cards left to connect decreases from 16 to 12, then 8, and lastly, 4. However, another way to solve this game, as seen on Connections Archive, is considered *hard* mode. In this version of the connections game, the user can submit all solved connections simultaneously, as seen in figure 6.8d. In this hard mode, if even one of the connection groupings is incorrect, the system logs it as a wrong answer, and then the user begins from the start with three remaining attempts.

The 16-card Connections game is designed to take about 5 minutes to play. However, for our user study, we wanted the task to take about 15-20 minutes to complete, allowing the users to utilize the new features in their analysis. After piloting the Connections game with 8 different users, we decided to make the following modifications to the game: 1) The game would be played in "hard" mode, requiring the user to submit their solution to the puzzle all at once; 2) 32 cards (or two connection games) were used for each task in the experiment, requiring the participants to solve eight, four card connections; 3) The users would be expected to create a label to describe the connections between each group of four documents, (i.e. *Bit of Vocal Fanfare*). In designing these modified Connections games, we aimed for similar difficulty, cross-puzzle connections (i.e., word/meaning connections exist between the two original 16-card puzzles that are combined), and one correct answer. We utilized the puzzle archive from Connections Archive to help curate the datasets, utilizing old puzzles/solutions as well as the feature to play "easy", "medium," and "hard" games to help curate similar difficulty tasks.

[2]https://www.nytimes.com/games/connections



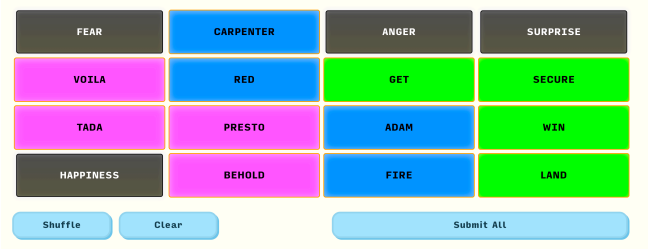
(a) Example Starting Board



(b) Example Board After Submitting One Connection



(c) Example Finished Board



(d) Example Board Where User Solves All Connections At Once Before Submitting

Figure 6.8: Connections Game Example

6.2.1 Task Comparison

While this task differs from those used in our prior work, it is still fundamentally a text-based analysis task. In this section, we compare the modified Connections game we developed to "Sign of the Crescent" to draw comparisons between the two as sensemaking tasks. We

break down our analysis based on the Pirolli & Card Sensemaking Process into two main flows: 1) Approximate Data Flow, which is represented as the rectangular boxes in the Sensemaking Model in figure 6.9 and 2) Process Flow, which are represented at the circles in the Sensemaking Model. The approximate data flow represents how information flows from raw data to the finalized reportable product during the sensemaking process, and the process flow represents the bottom-up and top-down processes the analysts work through to get the data into the presentable form. These stages are also organized by increasing effort and structure.

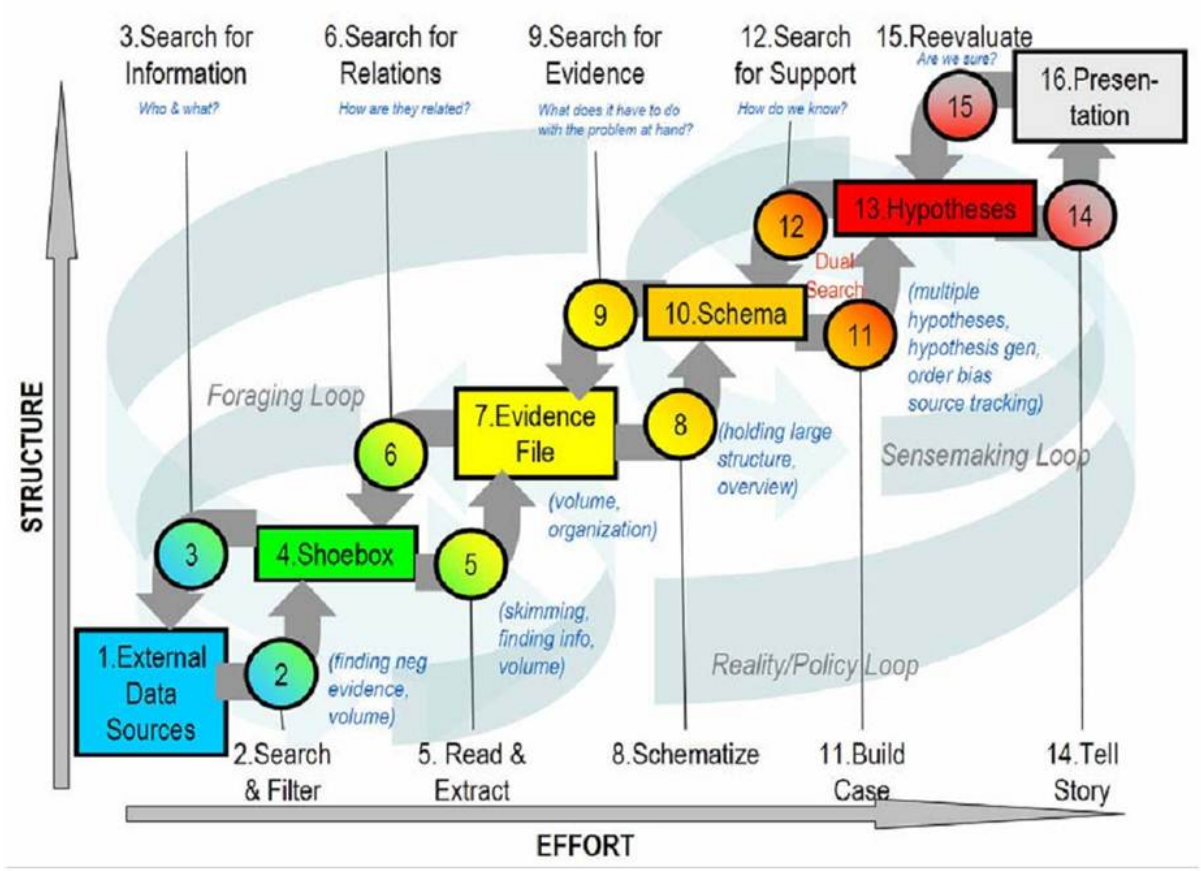


Figure 6.9: The Sensemaking Process as Defined by Pirolli & Card [73]. The process contains a series of steps used while an analyst develops a hypothesis or finding from unstructured data. Rectangular boxes represent the approximate data flow and the circles represent process flow during the sensemaking process.

Approximate Data Flow

External Data Source - This stage involves querying "Raw Text/Evidence," such as information from the web or a database. In our implementation of Sign of the Crescent, we do not support the external data source stage. In our version of Sign of the Crescent, we provide analysts with a smaller subset of documents that have already been determined to be relevant to their analysis. As a result, we do not utilize an external data source in our prior work with Sign of the Crescent. Similarly, in the modified Connections game, we do not support the external data source stage. In the modified Connections game, users are provided with a pre-determined set of cards for their analysis. This means that, like in Sign of the Crescent, the stage involving querying an external data source is not part of the process, as all necessary information is already included within the given set of materials.

Shoebox - is defined as a "smaller subset of the external data that is relevant for processing." In Sign of the Crescent, we support the shoebox stage by providing the carefully curated set of documents that the participant begins their analysis with. In our prior work with Sign of the Crescent, each document provided for analysis is part of this shoebox. Additionally, in the modified Connections game, we similarly provide the users with a predetermined set of cards for their analysis where each card is part of this shoebox.

Evidence Files - are defined as "snippets extracts from items in the shoebox." In Sign of the Crescent, each document provided can contain multiple snippets of information to be extracted. For example, document CIA_25 contains the following text: *Report Date 25 April, 2003. CIA [From Dutch security]:—– Analysis of radioactive particles found in a truck driven by Tawfiq al Adel and Saeed Khallad and in the storage unit rented by Saleh al Ahdal and Riduan bin Attash reveals them to be powdered cesium 137. One ounce of this substance could spread radioactive fallout over 60 city blocks.* Some example snippets that can be extracted are: 1) Tawfiq and Saeed are working together, 2) Radioactive material is present, 3) Radioactive material was transported from a storage unit to a new location, 4) etc. Similarly, snippets can be extracted from the Connections cards. In the connections task, definitions or features of the words can be extracted from use in the analysis process. For example, a Connections word SAFETY has multiple definitions: *1. The condition of being protected from or unlikely to cause danger, risk, or injury. "they should leave for their own safety." 2. A defensive back who normally is positioned well behind the line of scrimmage.* Additionally, the word SAFETY rhymes with other words, such as tasty or hasty. In the connections task, any of the features described above about the word could be a snippet extracted from the card to use in the analysis process.

Schemas - are defined as "re-representations of the evidence." In our previous work with Sign of the Crescent, the schemas are the organizations that the users make within the immersive space while completing their analysis process with the documents (e.g., timeline columns and clusters of related information). Similarly, in the modified Connections game, the schemas will be organizations that the users make within the immersive space while completing their analysis process with the cards. During analysis in both Sign of the Crescent

and Connections, the users are able to organize the information or schematize in space to assist them during their analysis process.

Hypotheses - are defined as "tentative representation of conclusions with supporting arguments." In our previous work with Sign of the Crescent, while the users are forming hypotheses of what is happening in the dataset during their analysis, the users are able to position the documents they are comparing/contrasting next to each other. Additionally, the users are able to organize information related to a specific topic close to each other. For instance, a participant may have a hypothesis that there is something going on in New York City, and they are evaluating their hypothesis by looking for support within the document set. In the modified Connections game, we also support the data flow of building hypotheses. When the user is solving the Connections game, the users can position cards near other cards that they believe this card could go with. For example, placing two cards, Saw and Comb, near each other. The user can then search for support for their hypothesis by trying to identify other cards that might belong with these connections, trying to find other words that are objects with Teeth (zipper or gear).

Presentation - is defined as the "work product" developed from the analysis. In our prior work with Sign of the Crescent, the participants were asked to write a report of their findings after completing their sensemaking process, which acts as the work product. In our modified Connections game, the work product is eight groups of four cards with a label generated by the user to describe each connection (i.e., the sorted groups with semantic labeling). This work product is a different presentation of findings than in our prior work but still fits the Pirolli & Card definition of the work produced developed from the analysis.

A table summarizing the experimental task comparison for the approximate data flow can be found in table 6.2.

Process Flow

In the model, there are two sets of processes described: Bottom-Up (from data to theory) or Top-Down (from theory to data). Here we break these two processes down and present how the Connections game provides us with many of the same sensemaking process flow stages as our prior work with Sign of the Crescent.

Bottom-Up Processes (From Data to Theory)

Search & Filter is the process where the external data sources (e.g., databases) "are queried by an analyst, and the results of those searches are filtered (judged) for relevance." The results from this stage are stored for later analysis in the Shoebox. In both Sign of the Crescent and Connections, we provide the participants with a shoebox of relevant data for their analysis; we do not support the Search & Filter stage in either of these tasks.

Read & Extract In this stage, the analyst "reads to extract nuggets of evidence that may

Approximate Data Flow Stage	Definition as defined by Pirolli & Card [73]	Represented in Sign of the Crescent?	Represented in Modified Connections Game?
External Data Source	"Raw Text/Evidence that can be queried"	No	No
Shoebox	"Smaller subset of the external data that is relevant for processing"	Yes	Yes
Evidence Files	"Snippets extracts from items in the shoebox"	Yes	Yes
Schemas	"Re-representations of the evidence"	Yes	Yes
Hypotheses	"Tentative representation of conclusions with supporting arguments"	Yes	Yes
Presentation	"Work product"	Yes	Yes

Table 6.2: Table summarizing the comparison between Sign of the Crescent and the Modified Connections Game when evaluated on the Approximate Data Flow stages of the Sensemaking Process.

be used to draw inferences or support or disconfirm theory." In Sign of the Crescent, as the users read the documents, they extract snippets from the files that may be relevant to the overall analysis process (e.g., Tawfiq and Saeed are working together, Radioactive material is present, Radioactive material was transported from a storage unit to a new location, etc. from CIA_25). Additionally, in the modified Connections game, the user is extracting ways in which the card they are looking at can be connected to other cards in the dataset through connections such as synonyms, features of the item, features of the word, or even word associations.

Schematize During this stage, the information can be "re-represented in some schematic way... Where the evidence may be organized into small-scale stories about typical topics or in answer to typical questions (e.g., who, what, when, where, why, how) that are used to organize raw evidence." One of the key features of IST is that it provides users with the ability to form meaningful organizations or schemas during an analysis process. Both of these tasks support the schematization process.

Build Case is where "a theory or case is built by additional evidence marshaling to support or disconfirm a hypothesis". In Sign of the Crescent, as the user identifies relationships and develops their hypothesis, they build a case to support this hypothesis through the information provided in the document set. Similarly, as the user finds connections within the Connection task, they are developing a hypothesis of which documents go together and building support for that hypothesis by identifying a strong group of 4 cards that match it.

Tell a Story is where a "presentation or publication of a case is made to some audience

(client)." In the Sign of the Crescent, the users have previously developed a written report of their findings, which acts as the presentation of their findings. In the modified Connections game, the presentation the user develops is the eight groups of 4 cards with a label describing the connections. Again, these types of presentations are different; however, both are an outcome of the analysis process that is submitted to the client (experimenter).

Top-Down Processes (From Theory to Data)

Reevaluate is the stage in which the "inquiries or feedback from clients of a presentation may generate re-evaluations of the current theory developed by an analyst," in which the analyst would need to search for additional support or evidence. In our prior user study designs and in the Connections user study design, we do not provide feedback to the users on the presentation (written report or organized connections). However, in both Sign of the Crescent and Connections, as the user is developing their presentation, they may need to reevaluate a theory or hypothesis after realizing they are missing a piece of evidence to support their claim, or in the connections game, the user has 3 cards grouped and is missing a last card to support that connection idea. In both cases, this could trigger the user to go back and reevaluate their hypothesis.

Search for Support is the process of reevaluating their hypothesis through "re-examining the lower-level schematic organization of basic facts." Sign of the Crescent and the modified Connections game within IST both support this stage when information is reorganized around the space to help the analysis in identifying support for their hypothesis.

Search for Evidence is the process that may require the user to re-examine the existing evidence or require the analyst to search for new evidence. In Sign of the Crescent, this process can be triggered by the user recalling back to which snippets were extracted from the documents or searching for new snippets in the documents that may help support their analysis. In the modified Connections game, if the user thinks they have identified a potential connection for a card (e.g., SAW is an object with teeth), they may search for evidence within the non-connected cards (other cards that are objects with teeth). Additionally, they may think about other alternative snippets that could be used to connect the word in the dataset (SAW as an action or tool).

Search for Relations is the process where the analyst inspects "Nuggets of information in an evidence file that may suggest new patterns (e.g., people linked to other people) that generate hypotheses about plausible relations among entities and events." In Sign of the Crescent, this process can be triggered when the user goes back to the bulletin board to try to identify new relationships based on the information they have already gathered in their analysis. In the modified Connections game, this process can also be triggered by the user going back to the bulletin board to try to identify different underlying connections or when a user thinks about a different way in which the words could be connected (e.g., secondary definition).

Search for Information In this process, the "new hypotheses generated from processes

Approximate Process Flow Stages	Process	Definitions as defined by Pirolli & Card [73]	Represented in Sign of the Crescent?	Represented in Modified Connections Game?
Search & Filter	Data to Theory (Bottom-Up)	"External data sources are queried by an analyst"	No	No
Read & Extract	Data to Theory (Bottom-Up)	"Reads to extract nuggets of evidence that may be used to draw inferences or disconfirm a theory"	Yes	Yes
Schematize	Data to Theory (Bottom-Up)	"Re-represented in some schematic way... Where the evidence may be organized into small-scale stories about typical topics or in answer to typical questions that are used to organize raw evidence"	Yes	Yes
Build Case	Data to Theory (Bottom-Up)	"A theory or case is built by additional evidence marshaling to support or disconfirm a hypothesis"	Yes	Yes
Tell a Story	Data to Theory (Bottom-Up)	"A presentation or publication of a case is made to some audience"	Yes	Yes
Reevaluate	Data to Theory (Bottom-Up)	"Inquiries or feedback from clients of a presentation may generate re-evaluations of the current theory developed by an analyst"	Yes	Yes
Search for Support	Theory to Data (Top-Down)	"re-examining the lower-level schematic organization of basic facts"	Yes	Yes
Search for Evidence	Theory to Data (Top-Down)	"re-examining the existing evidence or require the analyst to search for new evidence"	Yes	Yes
Search for Relations	Theory to Data (Top-Down)	"Nuggets of information in an evidence file that may suggest new patterns that generate hypotheses about plausible relations among entities and events"	Yes	Yes
Search for Information	Theory to Data (Top-Down)	"New hypotheses generated from processes at higher levels may cause the analyst to dig deeper in the raw data"	No	No

Table 6.3: Table summarizing the comparison between Sign of the Crescent and the Modifieid Connections Game when evaluated on the Approximate Process Flow stages of the Sensemaking Process.

at higher levels may cause the analyst to dig deeper in the raw data", which may cause them to run new queries to bring in new data. Again, since our prior Sign of the Crescent and modified Connections tasks provide the participants with a Shoebox, we do not see the process of Searching for Information within our experimental tasks.

A table summarizing the experimental task comparison for the approximate process flow can be found in table 6.3.

While Sign of the Crescent and the modified Connections games are different tasks, both are sensemaking tasks that share many similarities when we look step by step at the Pirolli & Card sensemaking model. The main difference between these two tasks when looking at the stages of the sensemaking process is around the Presentation (written report or sorted groupings of cards with labels). However, there are some limitations when it comes to utilizing the modified Connections games and changing the task. First, since we no longer focus on longer or more traditional text-based sensemaking tasks, we may lose some of the nuances in how the users organize their spaces (i.e., timelines, reporting agency grouping, etc.). Second, since this task is about grouping information into groups of 4, we may also lose some freedom in how the participants organize the information in the world around them. For this reason, we included the Freestyle IST condition in our user study so we could

make comparisons to a baseline IST where the task provided to the users was similar (i.e., a Connections-based task). Lastly, some of the new organizational features were initially designed for use with intelligence analysis text-based tasks, e.g., the timeline automated layout and discovery inking, which were designed to work on the metadata of the text-based documents. With this in mind, in our updated feature prototyping, we also redesigned some of the features to be more effective with the new connections-based task.

6.3 New Organizational Features Iterative Design

After considering the feedback we received during the expert feedback sessions and the modifications needed to support the Connections based tasks, we implemented the following changes to IST.

6.3.1 IST Document Modifications

The first set of modifications we made to our prototype surrounded the IST artifacts. We modified the documents to be multiple pages in length, as seen in figure 6.10. On the first page is the word for the connections game. On the next page, we present the top definition(s) of the words that were retrieved from the Oxford Language Dictionary. After the definitions, we presented a list of synonyms related to the word, retrieved from Thesaurus.com. After hitting the next page button again, the document returns to the title page. The pages were scrollable in either direction and when the definition/synonym list was open, the card's title was displayed above the card with a white background for readability purposes. We elected to provide the definitions and synonyms to the user to allow them to access alternative ways of linking information together for the connections game.

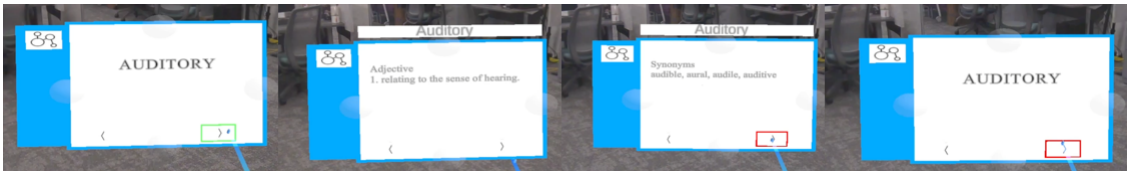


Figure 6.10: An example Card for the Connections Based Game. The document had two arrow buttons to access the word's definitions and synonyms. When the next page button is pressed, the card is scrolled through the definition and synonym lists. When the definitions or synonyms are displayed, a title bar appears over the document with the document title.

6.3.2 Clustering Modifications

The first modification we made to the clustering feature was adding anchor points to the clusters themselves, as seen in figure 6.11 left. Adding the anchor point to the cluster container allowed the user to link information to the cluster container or to link multiple clusters together. This feature was requested during our expert feedback session. The next modification we made was to support multiple color options for the clusters, as seen in figure 6.11 middle, again, as requested during the feedback sessions. For this study, we provided the users with 4 options for the clusters - green, yellow, red, and the default color blue. In creating the color-changing feature for this user study, we made a design decision to hide the automated layout options (timeline, column, and row) in place of the color-changing menu. This decision was made for two reasons. First, for simplicity. We did not want to overload the interface panel with too many buttons and wanted to keep training time for using the system low to prevent a long onboarding time. Secondly, we believed that with a 15-20 minute connections-based task, the participant might not utilize the automated layout features during their analysis. Additionally, the participants could always rearrange the information inside of the cluster if they wanted to tidy up the clusters. While we decided to choose the color-changing option over automated layouts for this user study, we believe that both features would be beneficial during sensemaking tasks to support the analysis process. The last modification we made to the clusters was the ability to place labels directly into the clusters, as seen in figure 6.11 right. This feature was suggested in the expert feedback sessions as well.

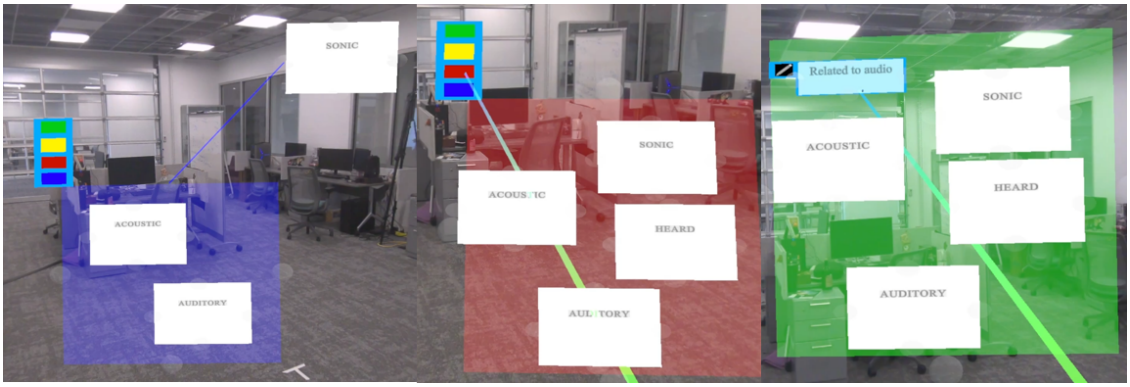


Figure 6.11: Modifications made to the clustering features. Left: Anchor points were added to the cluster container (top, right, bottom, left), which can be used to link information to clusters or link clusters to clusters. Middle: User Interface panel in the top left corner for changing the clusters' color with preset options of green, yellow, red, and blue. Right: Added the ability to place labels directly into clusters.

6.3.3 Manual Linking Modifications

We implemented two changes to the manual linking features. First, we developed a way for the participants to remove unwanted links, as seen in figure 6.12. Second, we added the ability to change the color of the links using a radial menu, as seen in figure 6.13. There are 4 color options for the linking features: green, yellow (default), red, and blue. We selected these colors based on the feedback we received in the expert feedback session surrounding the ability to express a level of confidence in the links, where P4 stated they could imagine using color to show that "green link means high confidence, red means explore more." In addition to the changes made to the linking features, we also increased the size of the colliders on the anchor points, so they were slightly larger than the spheres, to make them easier to select using the ray from a distance.



Figure 6.12: The process of deleting a link. First, the user hovers over the unwanted link using the ray, and then the user clicks the trigger on the controller, resulting in the link disappearing, right.

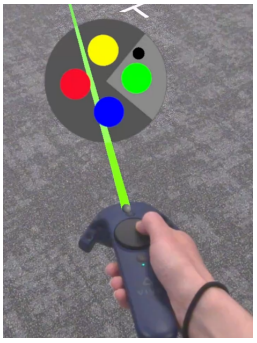


Figure 6.13: Radial Menu for changing the color of the links. The menu is opened when the user presses down on the controller trackpad. Then, the user can move their thumb around the track pad, which moves the cursor (black dot). Whenever the user releases their thumb, that color becomes the new link color. For example, if this user released their thumb, the new link color would be set to green.

6.3.4 Discovery Linking Modifications

We implemented one big change to the discovery linking feature. Previously, the user would select text from a document and press the discovery linking button on the document user

interface panel. This would then search the entire document set for any other document that contained that word and bring it around the spawning document. During the feedback sessions, we received a suggestion that we could utilize the metadata of the documents to make the results from the discovery linking feature more powerful. Building on the idea of using the metadata and trying to implement the discovery linking feature in a way that would work for the Connections tasks, we landed on a new iteration of this feature. For the new implementation, discovery links can be opened on any card without selecting any text, as seen in figure 6.14 left. After pressing the discovery linking button on a card, the back end of IST identifies any documents that have a matching word in the definition list or synonym list and will place those cards around the spawning document, as seen in figure 6.14, middle-left, where the keywords/synonym matches used word-stemming (e.g., running –> run or runs –> run) to ensure that matches are found regardless of a word's part of speech. Then, when the documents are displayed in the discovery link, a subtitle box is displayed below the found cards with a list of the matching keywords/synonyms as seen in figure 6.14 middle-left.

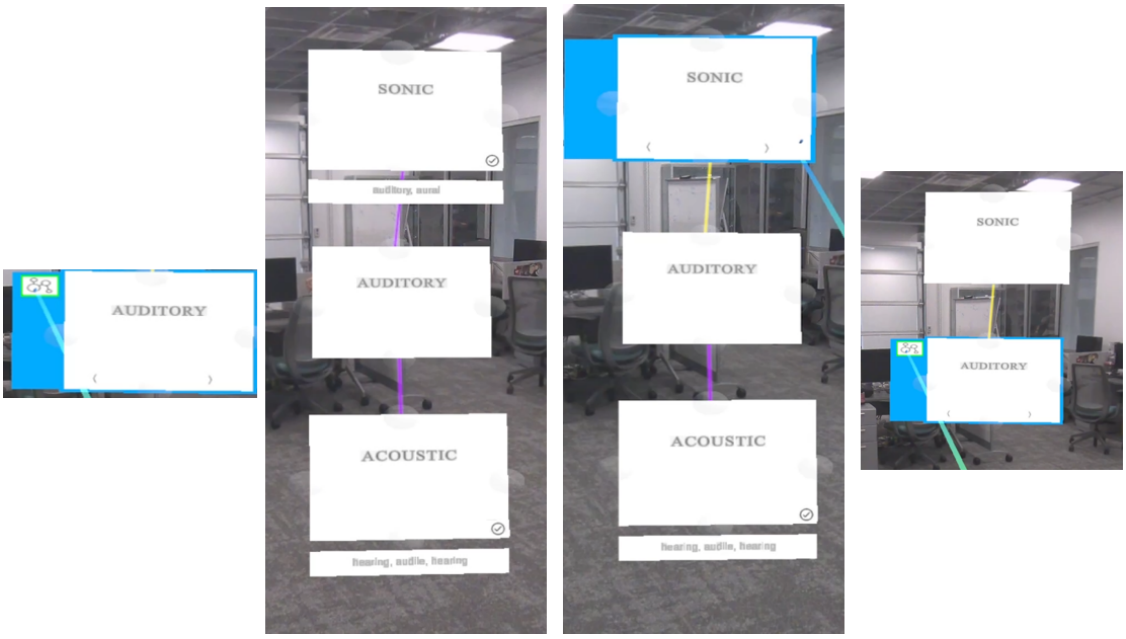


Figure 6.14: Discovery Linking Feature Example. Left - the button used to open a discovery link, Middle-Left - An open discovery link (purple links) on the work AUDITORY. Middle-Right - State after the check box (keep) button is pressed on SONIC. Right - After the discovery link is toggled off, SONIC is kept around, and ACOUSTIC is returned to its previous location.

Another modification we made to the discovery linking was around when the discovery linking button is visible to the participants. Since the discovery linking feature is a toggle (only one discovery can be opened at a time), we elected to hide the discovery linking button from all documents except the spawning document when a discovery link is open, as seen

on SONIC in figure 6.14 middle-right, ensuring that participants would need to close the previous discovery before opening a new one. The new implementation of the discovery link still supported the ability to keep a document from a discovery link, as seen in figure 6.14 middle. To keep a document, the check box is clicked on the bottom right corner of the document, for example, SONIC, which creates a new regular link between the documents, so when the discovery link is toggled off as seen in figure 6.14 right, the document stays around instead of going back to its last known location.

Another modification we made to the discovery linking feature was around when to move documents when a discovery link is opened. In the first iteration of discovery linking, we would move any document when a matching word was found if it was not already in a cluster. However, this meant that if a document was already linked with other documents, it would get moved during a discovery link. To keep document movement during discovery links consistent, we decided that if a document is already clustered or linked, it would not get moved during a discovery link. However, to ensure that the user still sees the connections, we still draw a link to the document's current location. The last modification we made to discovery linking was the color of the links themselves. Since we added color-changing options to the manual links (green, yellow, red, and blue), we wanted to ensure that discovery links were a different color than any of the manual link options. We selected purple as the new color for the discovery links.

6.3.5 Other Modifications for Experimental Task

Submitting Solution

For our Connections task, we added a button attached to the keyboard table that allowed the participants to submit their solution to the system, as seen in figure 6.15. When the submit button was pressed, a timestamp was saved to data analysis files, and all 3D information about document, link, and cluster locations was saved to a file for analysis.

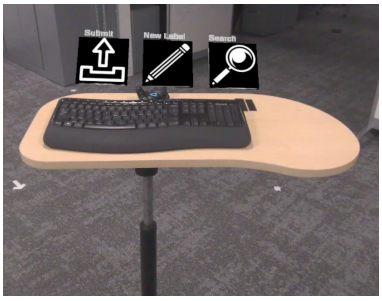


Figure 6.15: Keyboard Table used for Connections-Based task. The table had three buttons attached to it: Submit, New Label, and Search.

Disabling/Hiding Features Additionally, to support the experimental design, we needed a way to toggle groups of features on/off for the different conditions. For this, we implemented

some buttons within the Unity editor screen to be turned on/off to remove the clustering and linking features from the system. When Clustering was turned off, documents could not be overlapped to form a cluster. When Linking was turned off, the discovery link button was removed from the document user interface panel, and all of the anchor points were also turned off. This allowed us to switch which features were accessible to the user when running the user study. Lastly, we also disabled some of the prior IST features that were not needed for this study, including highlighting, copying/pasting to/from the clipboard, annotating documents, and note-taking.

While we implemented many changes to the features based on the expert feedback sessions to better support the new Connections-based task, we did not implement all the suggestions made by the experts for this user study. For instance, five of the experts suggested the ability to minimize or set custom sizing for the clusters. This feature could be very beneficial for larger document sets and prevent clutter within the system; however, within the Connections game, all of the cards are always visible to the user, so we decided to keep all documents visible to the user. Of course, there were other suggestions made by the experts that were not implemented in our iterative design as well. We felt that we implemented many first-round changes to the features and highlighted some of the other suggestions made in the results from the expert feedback sessions since these suggestions can provide the IA community with insights into other features/modifications that could be beneficial to explore further.

6.4 User Study

6.4.1 Goals and Research Question

This study aimed to understand how the addition of new organizational features within an immersive analytics system affects the sensemaking process.

The research questions were as follows:

RQ 3.1 - How do the new organizational features affect the spatial structures created during sensemaking?

Our previous work evaluated the types of high-level organizational structures used within IST. Additionally, we have looked at lower-level, deeper organizational features. We want to understand how these new organizational features affect the schemas created during sensemaking, whether they change the overall space usage, and whether they lead to more organized spaces.

RQ 3.2 - How do the new organizational features affect the strategies used for sensemaking?

In previous work, we have detected high-level strategies utilized for sensemaking within IST.

Since these new features were focused on assisting with schema formation and foraging within IST, we want to understand how these new features affect the overall sensemaking strategies used.

RQ 3.3 - How do the new organizational features affect cognitive load while completing a sensemaking task?

This research aims to assist the user in developing additional ways for analysts to offload their cognition into the environment. To measure the effects of these features, we want to understand how these features affect cognitive workload while completing analysis.

RQ 3.4 - How do the new organizational features affect the system's overall user experience?

In addition to understanding how these features affect the sensemaking process, organizations, and user workload, we also want to understand if these new features affect the overall system usability.

RQ 3.5 - How do the new organizational features affect task performance?

Lastly, we want to understand if the features affect the overall task performance, such as time for task completion and overall correctness of analysis based on the ground truth.

6.4.2 Apparatus

For this study, we used a Varjo XR-3 [3] head-worn display in augmented reality passthrough mode. The Varjo-XR 3 has a focus area of 1920 x 1920 pixels per eye, a peripheral area of 2880 x 2720 pixels per eye, a refresh rate of 90 Hz, and a 115-degree field of view. The headset ran on a desktop PC with an AMD Ryzen 7 5800X 8-Core processor and a NVIDIA GeForce RTX 3080 graphics card. The tracked area for the study was a 4m X 2.9m area, in which the participants could freely walk around while completing their analysis tasks. The participants were trained on the size of the tracked space during the tutorial phase of the study. Since the Varjo-XR 3 is a tethered headset, the experimenter watched the participant during the session to ensure the participant did not end up tangled in the headset cord. For interacting with the system, participants used a VIVE 2 Controller [4] in their dominant hand. The 3D interaction technique used in this IST prototype was the Hand-Centered Object Manipulation Extending Ray-casting technique (HOMER) [11]. The participants used a wheeled table with a wireless keyboard for text entry. The keyboard-table had a VIVE 1.0 tracker [5]. The tracked area was accomplished using the SteamVR 2.0 Lighthouse Tracking system. The IST prototype was built in Unity version 2020.3.9f1. The documents were 0.5 x 0.3 meters with the text size in Text-Mesh Pro set to 60.

[3] https://varjo.com/products/varjo-xr-3/
[4] https://www.vive.com/us/accessory/controller/
[5] https://business.vive.com/eu/product/vive-tracker/

6.4.3 Experimental Task and Conditions

As presented in section 6.2, the experimental task for this user study was based on the NYT Game called Connections. For the study, we developed four modified Connections-based tasks, which consisted of 32 cards (with their definitions and synonyms), resulting in four experimental task datasets. Additionally, we developed a training dataset that consisted of 16 cards (with definitions and synonyms). As with our prior user studies, the cards started loaded onto a virtual bulletin board in a randomized order. The order in which the tasks were presented to the user was rotated using a Latin square design.

To understand how each of these new features affects the different measures we are interested in, we wanted to evaluate how each of the features affects the overall sensemaking in a manner where we could make direct comparisons among the features. To be able to make these comparisons, we conducted a within-subjects user study design with four conditions. The four conditions were: **Freestyle IST (Baseline)** - IST with no new features, **IST Clustering** - IST new clustering related features, **IST Linking** - IST plus the linking features (both manual and discovery), and **IST Clustering & Linking** - IST with the new clustering & linking features. For each condition, features that were not available during that condition were hidden from the system so the user would not have access to them during their analysis. Additionally, participants were always able to access definitions/synonyms of the words and execute searches within the system. The order in which the participants experienced the conditions was counterbalanced using a Latin square design.

6.4.4 Participants

We recruited 39 participants who were affiliates of Virginia Tech, were native English speakers, were at least 18 years of age, and had normal or corrected to normal (with glasses or contacts) vision. Due to scheduling conflicts and data collection challenges, 7 participants' data was removed from the user study. This resulted in 32 total participants (15 female, 17 male), with a mean age of 20.875 (SD = 2.01, median = 20.5). The participants had varying levels of VR/AR experiences, including four participants with no prior experience, 14 participants with one or two prior experiences, eight participants with 3-10 prior experiences, and six participants with 10+ prior experiences. Additionally, the participants had varying familiarity with the experimental task game, with 11 participants having no prior experience with the NYT Connections Game, 4 participants who had played once or twice before, 4 participants who play once or twice a month, 4 participants who play once or twice a week, 6 participants who play close to daily, and 3 participants who play daily (21 total with prior experience). The study was approved through the Virginia Tech Institutional Review Board.

6.4.5 Measures

We collected data for our analysis in a number of ways. First, we administered a pre-study background questionnaire that gathered background information on the participants, such as age, gender, VR/AR experience, and Connections game experience. During each experimental task, we collected save files and log files. The save files were generated every 15 seconds and captured all IST artifacts (cards, labels, clusters, and links) locations during that timestep (x,y,z, yaw, pitch, and roll). The log files capture timestamps of all participant actions within IST, including headset/controller/keyboard location, document movement, and any feature interactions (e.g., new label, new link, new cluster, etc.). During each experimental task, we gathered screen recordings from the participants' first-person viewpoint. After each experimental task, the participants completed a series of questionnaires, including the NASA Task Load Index (TLX) [42], to measure cognitive demand, the User-Experience Questionnaire (UEQ) [55] aimed to gather feedback on the user's experience in each condition, and a series of custom Likert-scale (Strongly-Disagree through Strongly Agree) questions: 1) I feel confident in my solution to this task. 2) This task was easy to complete. 3) The features provided in this condition supported my sensemaking in completing the task. All Likert scales in the questionnaires were on a 7-point scale. Additionally, after each task, we conducted a semi-structured interview with the participants about how they used the features in the condition and their overall analysis strategy. After the participants had completed all 4 experimental tasks, we conducted a post-study interview to gather insights into what the participants thought about the different features, which feature was most helpful to them in completing the analysis task, which feature was most helpful to them in staying organized during their analysis, which condition best supported their overall strategy for solving the Connections-based game, and suggestions/comments on the features as whole.

6.4.6 Procedure

The user study was designed to take about 2.5 hours to complete. The user study session started with welcoming participants to the lab, providing the user with a quick background on the user study and the formal consent process. After consenting to participate in the research, the participants completed the background questionnaire. Then, we began the tutorial phase of the study. First, we went over a PowerPoint presentation with the participants that provided them with background on IST and the different conditions they would be experiencing during the user study and training on how to play the Connections-based task. During the Connections task training, we guided them through solving a connections game, provided them with helpful tips, and trained them on how to generate the labels for submitting their solution.

Next, we trained the participants on all of the hardware they would be using during the experiments before they donned the HWD for the in-headset tutorial phase. We provided the

participants with the 16-card connections training dataset for the tutorial. Before beginning the training on how to use the different features of IST, we trained the user on the tracked area and wheeled keyboard table so they could feel comfortable walking around in the space during their analysis. For the tutorial, we had the participants solve one of each of the connections in the tutorial dataset using one condition's set of features. To keep the training phase consistent, the participants all learned the features in the following order: Freestyle IST, Clustering IST, Linking IST, and lastly, Clustering & Linking IST. After having the participant solve each connection, the participants were guided through how to create a label for their connection and asked if they had questions relating to that set of features. At the end of the tutorial phase, the participants would have four groups of four cards with a label describing the connections. At this point in the analysis, the participants were then trained on how to submit their solution to the system using the button on the tracked keyboard and table. After completing the tutorial, the participants were asked to take the headset off while the experimenter prepared the first experimental task for the participant.

Before starting the experimental task phase of the user study, the participants were provided with some additional details on the tasks:

First, we reminded the participants that in the connections game, a card could appear to be related to multiple categories and that this could be an opportunity for them to utilize some of the different features to represent the possibility of where a card may belong.

Second, we reminded the participants that they would be solving a 32-card version of the game, which was designed to take about 15 minutes to complete. However, they would have a buffer of 5 additional minutes as needed to get their solution submitted. The users were informed that the goal was to complete the task as accurately as possible in the time provided and reminded that they have definitions/synonyms of the words, the search feature, and any features related to the condition they were experiencing to help them in they got stuck during their analysis. We gave time warnings at 5 minutes, 2 minutes, and 1 minute remaining, then at the 15-minute mark, we moved into the buffer period where the participants were told to form a best-educated guess on which cards belong together and create a label to get the solution submitted since we were requiring complete solutions to hit the submit button.

Lastly, we reminded the participants that the way in which they organized the immersive space was up to them, as well as the way they utilized the features of the different conditions. However, we encouraged them to utilize all of the different conditions' features when they had them at their disposal.

Before starting the experimental task, the participants were informed which condition they were going to experience and provided with a quick reminder of the features of that condition. Then, the participant would don the HWD again and begin the experimental task. After submitting each of the tasks, the participants took the HWD off, and took a seat while completing the post-task questionnaires and post-task interviews. Then, we would repeat the process until the participant had completed all four experimental tasks, questionnaires, and post-task interviews. After completing the experimental task phase of the user study,

the participants responded to semi-structured interview questions related to the features. The completion of the post-study interview marked the completion of the study.

6.4.7 Results & Discussion

This study utilized various statistical methods to ensure robust and accurate data analysis. For data that did not adhere to a normal distribution, we employed the Aligned Rank Transform (ART) repeated measures ANOVA [94]. This non-parametric technique is effective in handling non-normally distributed data by aligning and ranking the data before applying ANOVAs, thus maintaining the integrity of the non-parametric data characteristics. For post hoc analysis, we used the ART-C approach for pairwise comparisons [23], with Tukey's HSD correction to control for Type I error. When running our analyses, we included a random effect for participant ID to account for individual-level variability.

For the analysis of Likert scale data, which is ordinal and often does not meet normality assumptions, we also utilized ART ANOVAs with ART-C for post hoc analysis. In instances where non-parametric data required pairwise comparisons, we applied the Wilcoxon Signed Rank-Test. This test is suitable for comparing paired samples that are not normally distributed.

Throughout the analysis, we evaluated the effect of the experimental condition (Freestyle, Clustering Only, Linking Only, and Clustering & Linking) and two-by-two factors: the Presence of Clustering (No vs. Yes) and the Presence of Linking (No vs. Yes). This thorough evaluation allowed us to understand how these new features affected different aspects of the sensemaking process and overall system usability.

RQ 3.1: *How do the new organizational features affect the spatial structures created during sensemaking?*

Angular Space Usage

One feature of the analysis spaces we were interested in evaluating was whether the new organizational features affected the total angular degrees used for document layout. To look at the overall angular space usage, we centered the spatial layouts based on the participants' most dwelled position (the participants' center of the space). We then calculated the angle of the longest arc (i.e., the 'leftmost' point to the 'rightmost' point from the user's center point while looking at only the x and z axes) in each of the spatial layouts, which represented the total degrees used for the document layout. The average angular space used across conditions was Freestyle = 184.68° (sd = 43.31), Clustering Only = 192.74° (sd = 50.66), Linking Only = 200.53° (sd = 49.51), and Clustering & Linking = 188.81° (sd = 46.48). We found no effect of Condition ($F(3,84) = 1.44906$, $p = 0.23439$, $\eta_p^2 = 0.05$), Dataset ($F(3,84) = 0.30176$, $p = 0.82403$, $\eta_p^2 = 0.01$), or Condition*Dataset ($F(9,85) = 1.16824$, $p = 0.32570$,

$\eta_p^2 = 0.11$) on the total amount of angular space used for document layout.

Total Distances Between Documents Normalized by the Number of Documents (Overall Spread of Documents)

We found an effect of experimental Condition ($F(3,84) = 7.90$, $p = 0.0001$, $\eta_p^2 = 0.22$) on the total distances between the documents. Contrasts using ART-C revealed that *Freestyle* had less total distance than *Clustering Only* ($p = 0.0345$) and *Clustering & Linking* ($p = 0.0138$), as well as *Linking Only* having less distances than *Clustering Only* ($p = 0.0025$) and *Clustering & Linking* ($p = 0.0008$). Dataset and Condition*Dataset had no effect on the total distances between documents. The results of the normalized total distances between documents can be seen in figure 6.16.

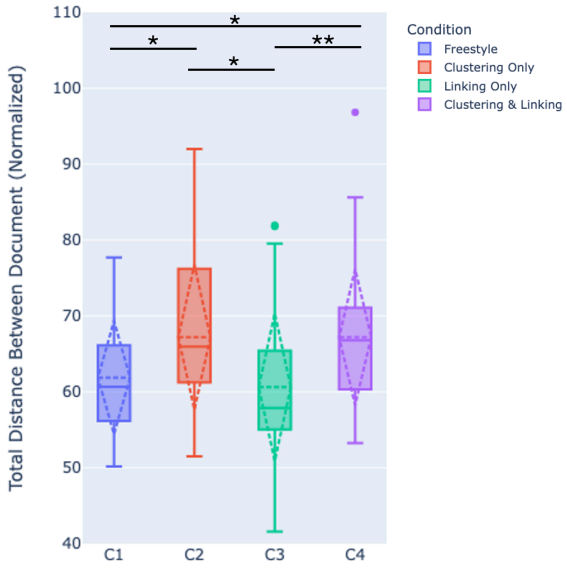


Figure 6.16: Breakdown of the (Total Distances Between Document / Number of Document). The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05, ** ≤ 0.001

Workspace Neatness

Another measure we were interested in evaluating is the overall neatness of the space using quantitative metrics such as Silhouette Score [76] and Calinski Harabasz (CH) [13] Scores. Both of these metrics are utilized to evaluate the goodness of clustering. Silhouette scores range from -1.0 to 1.0 and indicate how close a point or document is to its own cluster (cohesion) compared to the other clusters (separation). CH is also known as the variance ratio criterion and is an indicator for the sum of between-cluster variance and within-cluster variance for each cluster. With CH, the higher the score, the better the overall clustering, and for silhouette scores, closer to 1.0 indicates a very *strong* clustering or a more *tidy* space.

Prior work by Tahmid et al. [88], which evaluated different clustering techniques within IST, found that clusters formed without explicit clustering methods (i.e., Freestyle) had higher silhouette scores than conditions that had explicit clustering methods (e.g., our clustering only condition). Based on these findings, we expected to see similar results within our study, where conditions with explicit clustering techniques, *Clustering Only and Clustering & Linking*, would have a lower silhouette and CH scores than Freestyle and Linking Only.

Our analysis revealed an effect of experimental condition ($F(3,84) = 14.03$, $p < 0.001$, $\eta_p^2 = 0.33$) on silhouette scores. There was no main effect of Dataset or Condition*Dataset on silhouette scores. We also looked to see if there was an effect of the presence of Linking ($F(1,93) = 26.56$, $p < 0.001$, $\eta_p^2 = 0.22$) or Clustering ($F(1,93) = 23.84$, $p < 0.001$, $\eta_p^2 = 0.20$) on silhouette scores, revealing that both had an effect on the silhouette scores. We found similar results with CH scores, with Condition ($F(3,93) = 17.75$, $p < 0.001$, $\eta_p^2 = 0.36$), presence of Linking ($F(1,93) = 22.99$, $p < 0.001$, $\eta_p^2 = 0.20$), and presence of Clustering ($F(1,93) = 24.12$, $p < 0.001$, $\eta_p^2 = 0.21$) all having an effect.

Contrast analysis with ART-C revealed that *Freestyle* had higher silhouette ($p < 0.001$) and CH ($p < 0.001$) scores than *Linking Only* . *Clustering Only* had higher scores on silhouette ($p < 0.001$) and CH ($p < 0.001$) than *Linking Only* , and again, *Clustering & Linking* had higher silhouette ($p < 0.001$) and CH ($p < 0.001$) scores than *Linking Only* . We also evaluated the combined factors of the presence of Linking and Clustering, which revealed that when clustering is present, the silhouette and CH scores are higher than when clustering is not present. Additionally, we found that the silhouette and CH scores were lower than when linking was not present. The results of the silhouette and CH scores can be seen in figures 6.17 and 6.18, respectively.

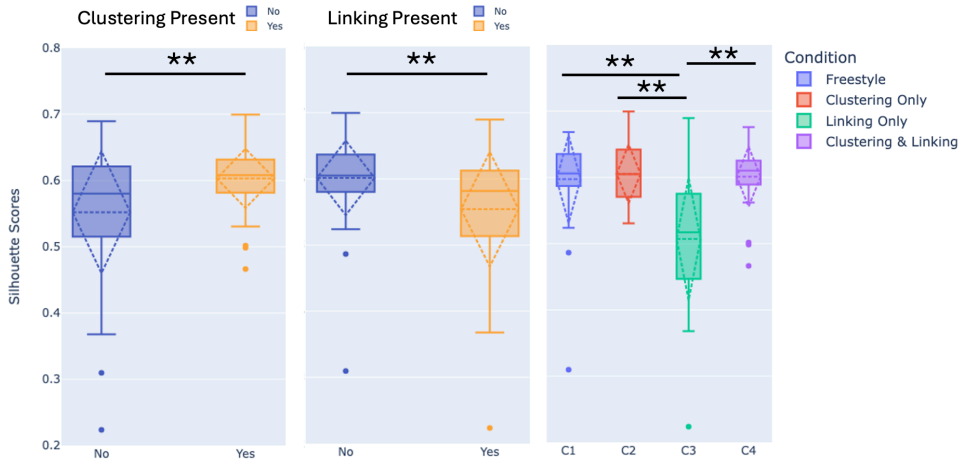


Figure 6.17: Breakdown of the Silhouette Scores. The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05, ** ≤ 0.001

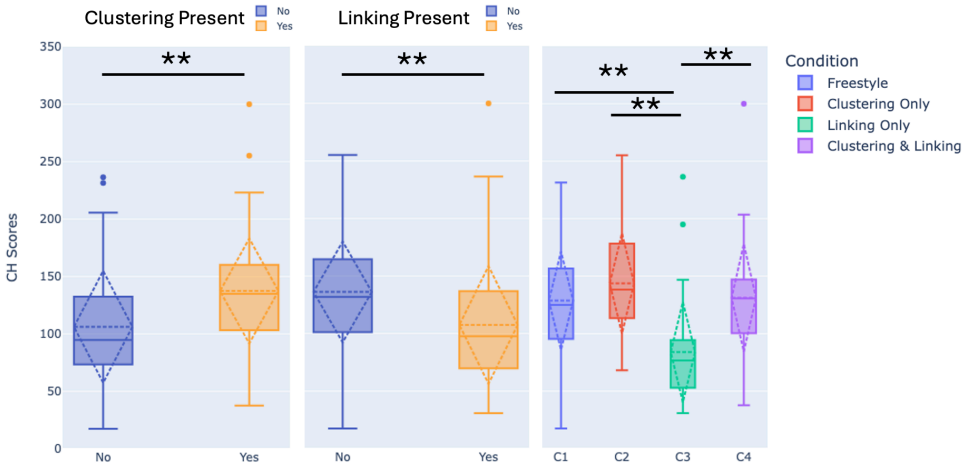


Figure 6.18: Breakdown of the CH Scores. The dotted line and diamond visualize the mean and standard deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05, ** ≤ 0.001

RQ 3.1 Discussion

In RQ 3.1, we asked how the new organizational features affected the spatial layouts created during the sensemaking task. Our analysis revealed that the total distances between documents (normalized) were significantly different between *Freestyle* and *Clustering Only* and again between *Freestyle* and *Clustering & Linking* . Additionally, we found that *Linking Only* was significantly different from *Clustering Only* and *Clustering & Linking* . One possible reason why the clustering features had higher total distance between documents could be that clusters provided users with an easy way to distinguish which group the documents were in (since they were all part of the same colored group), which meant that the documents did not need to be located as close to each other to show the distinction between the connections. Additionally, the clustering feature helps reorganize the information in the clusters to prevent overlapping documents, which could have led to the documents being more spread out than in the non clustering conditions. Another interpretation of this result is that in the clustering condition, the participants were able to easily push the connection further away from them to get it out of their *working area*. For instance, once the participants completed making a connection and felt confident about it, they could push the whole group away from them to clear up more working space. This could have led to a larger total distance between the documents (normalized).

We also evaluated the overall workspace neatness across conditions along with the total distances using Silhouette and CH scores. Based on prior results [88], we expected to see that the explicit clustering techniques would have lower workspace neatness than the non-clustering conditions, as multiple clusters can be placed closer to each other while still

providing users with a visual indicator that the clusters are different groups of information. However, our findings did not support this, our analysis revealing that *Linking Only* had worse overall workspace neatness than *Freestyle* , *Clustering Only* , and *Clustering & Linking* , and that there was no significant difference between *Freestyle* , *Clustering Only* , and *Clustering & Linking* . The feature design could be one possible reason why the *Linking Only* had the lowest Silhouette and CH scores. The Linking feature allows the participants to visually connect items in the space that are not located physically close to each other to show the relationship. In other words, the linking feature acts as an alternative to using space to show relationships within the analysis space. Whereas, in the conditions where clustering was present, the documents needed to be located physically close to be in the same cluster bin. Moreover, in the *Freestyle* condition, since there were no visual ways of connecting the information, the participants would use the document proximity to each other to make distinct groups in the analysis space.

RQ 3.2: ***How do the new organizational features affect the strategies used for sensemaking?***

Overall Clustering and Linking Feature Usage

One aspect of the new organizational features we aimed to understand was how the participants used these features during this sensemaking task to support their analysis process. During the tutorial phase of the user study, the participants were instructed that they could use the features of each condition however they wanted to, but they were encouraged to use the features when they had them at their disposal. Based on how we designed these features, we expected to see some overall trends in how the participants would use them during their analysis. For instance, we expected to see the participants use the clusters as a way to form the 8 groups of 4 documents. Therefore, we expected to see 8 total clusters in the conditions where the participants had the clustering features (i.e., Clustering Only and Clustering & Linking Conditions). As for linking, we expected to see the participants utilize the linking features as a way to show a possible connection between two words within the dataset. Therefore, we expected to see fewer links used during analysis and the 8 groups of 4 connections being formed with a few links and mostly by placing the documents in close proximity to each other, such as in the freestyle condition.

Another feature that we expected to see the participants utilize to represent uncertainty in their analysis process was the use of different colors. During the expert feedback sessions, one of the suggestions we received was to provide more color options to allow the analysts to represent overall confidence in their analysis. For this reason, we selected to use Yellow, Green, Red, and Blue as the colors for the participants to utilize in both the cluster colors and link colors. We expected to see the participants use the color of the clusters to indicate overall confidence in the clusters that they had formed. For example, using green to show high confidence in a grouping, and red to show uncertainty about the grouping.

Additionally, we expected to see the participants utilize different color links to represent the possible ways of connecting documents. For example, the word GAS could be linked to words like WATER, ELECTRIC, and CABLE as Types of Utilities, or GAS could be linked to words like PLASMA, LIQUID, and SOLID as States of Matter. In this case, different color links could be used to tie GAS to these different groups and show a potential way of connecting the documents.

Here, we present some trends identified in the use of Clustering Features and Linking Features.

Clustering First, we looked at the average number of clusters used for the participants' solutions in *Clustering Only* and *Clustering & Linking*. In *Clustering Only*, participants had an average of 7.875 (sd = 0.707, median = 8) clusters, and in the *Clustering & Linking*, participants had an average of 7.438 (sd = 1.413, median = 8) clusters. In the *Clustering Only* condition, 14/32 participants utilized at least two different color clusters. Again, in the *Clustering & Linking*, 14/32 participants utilized at least two different color clusters. In *Clustering Only*, the most common color used for clusters was the default color blue, with 203 clusters in total. The next most common color was green, with 27 total green clusters, then 14 total yellow clusters, and 8 total red clusters across all participants. In *Clustering & Linking*, again, blue (default) was the most common color, with 177 total clusters across the participants. Then green with 25 total clusters, red with 19 total clusters, and 17 total yellow clusters. We also looked at the average number of times clusters were made across the whole analysis session, with an average of 12.81 (sd = 2.76) in *Clustering Only* and 13.25 (sd = 3.53) in *Clustering & Linking*. Lastly, we were interested in when clusters were removed across the whole analysis session, with an average of 4.44 (sd = 2.30) in *Clustering Only* and 5.41 (sd = 2.84) clusters removed in *Clustering & Linking*.

Linking In *Linking Only*, participants utilized an average of 31.23 (sd = 13.31, median = 35.5) links for their solutions. In *Clustering & Linking*, participants utilized an average of 6.84 (sd = 7.87, median = 5) links for their solutions. In *Linking Only*, 11/32 participants used at least two different color links in their solution to the task. In *Clustering & Linking*, 3/32 participants used at least two different color links in their solutions to the task. In *Linking Only*, the most common color used was yellow (default) with 642 total links, then red with 136, blue with 123, and green with 98, across all the participants. In *Clustering & Linking*, the most common color was yellow (default) with 174 links, then green with 28 links, red with 13 links, and blue with 4 links across all the participants. We also looked at the average number of links created across the whole analysis, with 32.875 (sd = 12.64) in *Linking Only* and 7.28 (sd = 8.37) in *Clustering & Linking*. Lastly, we were interested in when links were deleted across the whole analysis, with an average of 5.53 (sd = 5.61) links deleted in *Linking Only* and 2.28 (sd = 2.54) in *Clustering & Linking*.

Discovery Linking For discovery linking, we wanted to understand how many times the participants opened a *discovery graph* and how many times that led to finding a hidden connection in the dataset (through using the keep button). Our analysis revealed that in

Linking Only, participants opened an average of 8.21 (sd = 5.72) discovery graphs, and in *Clustering & Linking*, the participants opened an average of 7.75 (sd = 6.74) discovery graphs. In total, 263 discovery graphs were opened in *Linking Only*, and 248 graphs were opened in *Clustering & Linking*. In *Linking Only*, the keep button from the discovery graph was used an average of 2.5 times (sd = 3.17), and in *Clustering & Linking*, the button was used an average of 1.88 times (sd = 2.92). Across all participants, the keep button was used a total of 74 times in *Linking Only* and 60 times in *Clustering & Linking*. While this metric does provide us with some insight into how many times the discovery linking feature aided the participants in finding a hidden connection, this is, of course, not representative of every time the feature helped uncover a connection.

System-Level Interactions

To examine the overall strategies for sensemaking across the different conditions, we began by examining the log files generated during the user study tasks. The log files collect each interaction within the system and a timestamp of when that interaction occurs. Then, using these log files, we can examine the frequency of each interaction across the different conditions. We first began by evaluating the interactions that the participants could use in each of the conditions, which included document grab information, definitions/synonyms access through the *next page button*, and searching for keywords/phrases.

Our analysis revealed a main effect of the experimental condition of a number of **Artifact Grabs** within IST ($F(3,84) = 4.18$, $p = 0.008$, $\eta_p^2 = 0.13$). The mean number of grabs for each condition were as follows: *Freestyle* = 319.81 (sd = 100.83), *Clustering Only* = 333.75 (sd = 85.18), *Linking Only* = 360.125 (sd = 108.59), *Clustering & Linking* = 370.31 (sd = 99.51). Post hoc analysis revealed that *Freestyle* required less grabs than *Clustering & Linking* (p=0.0088). There was no main effect of Dataset or Condition*Dataset on the number of grabs. The results of the artifact grab analysis can be seen in figure 6.19, left.

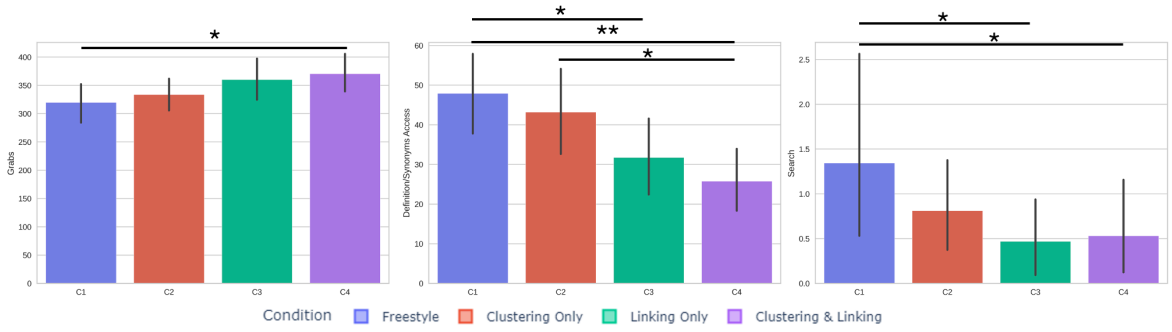


Figure 6.19: Average Total Number of Grabs Across Participants (left), Definition/Synonym Access (middle), and Search (right) interactions by condition. The vertical black line annotation shows the 95% confidence interval. Significance is denoted with * ≤ 0.05, ** ≤ 0.001

Our analysis revealed a main effect of experimental conditions on the number of **Definition/Synonym Access** interactions ($F(3,84) = 9.65$, $p < 0.001$, $\eta_p^2 = 0.26$). The mean count of interactions per condition was as follows: *Freestyle* = 47.90 (sd = 30.024), *Clustering Only* = 43.19 (sd = 31.73), *Linking Only* = 31.75 (sd = 28.09), and *Clustering & Linking* = 25.75 (sd = 22.92). Post hoc analysis revealed that *Freestyle* had a higher interaction count than *Linking Only* ($p = 0.006114$), *Freestyle* had a higher interaction count than *Clustering & Linking* ($p < 0.001$), and *Clustering Only* had a higher interaction count than *Clustering & Linking* ($p = 0.001$). *Clustering Only* was close to significantly higher interaction than *Linking Only* as well ($p = 0.065$). There was no main effect of Dataset or Condition*Dataset on the definition/synonym access. The Definition/Synonym access analysis results can be seen in figure 6.19, middle.

Lastly, we looked at the total **Number of Searches**. We found a main effect of Condition ($F(3,84) = 4.90$, $p = 0.003$, $\eta_p^2 = 0.15$), and a close to a significant effect of Dataset ($F(3,84) = 2.50$, $p = 0.065$, $\eta_p^2 = 0.08$). There was no effect of Condition*Dataset. The average number of searches per condition was as follows: *Freestyle* = 1.34 (sd = 2.902), *Clustering Only* = 0.813 (sd = 1.424), *Linking Only* = 0.468 (sd = 1.24), *Clustering & Linking* = 0.531 (sd = 1.67). Post hoc analysis revealed that *Freestyle* had more searches than *Linking Only* ($p = 0.004$), and *Freestyle* had more searches than *Clustering & Linking* ($p = 0.0138$). We also ran a post hoc analysis looking specifically at the experimental task dataset, and experimental task dataset two had fewer searches than experimental task dataset four ($p = 0.039$). The results of the number of searches analysis can be seen in figure 6.19, right.

In addition to examining the interactions present across conditions, we aimed to understand how the presence of clustering features affected the use of linking features and vice versa. Specifically, we analyzed system-level interactions pertinent to the clustering and linking conditions, including link creation/deletion, creating/merging/removing clusters, discovery linking usage, adding to a cluster, and all cluster and link actions combined.

First, we evaluated the effect of the presence of clustering on linking interactions. We used the Wilcoxon Signed-Rank test and found a significant effect of clustering presence on **Link Creation** (χ^2 (1) = 5.34, $p = 0.02$), indicating that fewer links were created when clustering was present. There were no significant effects of clustering presence on discovery link button usage or link deletion. Additionally, when aggregating all linking features, we found a significant effect of clustering presence on **Overall Linking Feature Usage** (χ^2 (1) = 4.05, $p = 0.04$), with fewer linking interactions observed when clustering was present.

Lastly, we evaluated the effect of linking features' presence on clustering feature interactions. Using the Wilcoxon Signed-Rank test, we found no significant effect of linking presence on clustering interactions, including creating a cluster, removing a cluster, merging clusters, adding to clusters, or all clustering features combined.

Physical Navigation Our prior work within IST has evaluated the total amount of physical space navigated during the analysis sessions. For our analysis, we looked at navigation breakdown across the experimental conditions. We found that participants navigated an average of 62.73 m (sd = 5.17) in *Freestyle*, 63.41 m (sd = 20.26) in *Clustering Only*, 65.84 m (sd = 21.81) in *Linking Only*, and 65.21 m (sd = 22.06) in *Clustering & Linking*. Our analysis revealed no effect of Experimental condition, Presence of Linking, or Presence of Clustering on the total amount of physical space navigated during the analysis session.

RQ 3.2 Discussion

RQ 3.2 asked how the new organizational features affected the strategies used for the sense-making task. In our analysis, we looked at the overall usage of the different organizational features and system-level interactions to explore this question. For the clustering features, the participants used these features as expected. Overall, the participants used the clusters to form each of the connections during the experimental task. The color-changing features on the clusters were not used that often. However, the participants who utilized the color options enjoyed that they could use them to help organize the connections they were confident in and which ones they wanted to come back to.

The linking features in *Linking Only* were not used as we expected them to be. We did not expect to see the participants create as many links as they did in this condition, where with an average of 31.23 links, the participants were effectively connecting all the items within a connection with a link. During the tutorial and prior to starting the experimental tasks, the participants were reminded they could utilize these features however they wanted during the analysis. However, maybe the participants felt the need to connect all of the information using the links to make distinct groups (as they were able to do in the clustering conditions). In *Clustering & Linking*, the participants utilized the links, as we were expecting to see the links utilized - to show a possible connection between items within the space before making a final decision on the document's placement. The color-changing feature was utilized more with the linking feature than with the clustering feature.

In looking at the system-level interactions, we identified significantly fewer grabs in *Freestyle* than in *Clustering & Linking*. We do not have a good interpretation of why we found these results. However, it could be possible that the participants moved the artifacts around more during their analysis because there were more features to interact with in *Clustering & Linking*, which made the analysis more engaging and led to them interacting with the information more during analysis.

As for Definition/Synonym access and the number of searches executed, we found that the participants accessed the definitions significantly less in *Linking Only* than in *Freestyle*. Additionally, the participants accessed the definitions/synonyms less in *Clustering & Linking* than in *Clustering Only* and *Freestyle*. Along with this, we saw fewer manual searches in *Linking Only* and *Clustering & Linking* than we did in *Freestyle*. A possible explanation

for why we saw these trends within the interactions is the presence of the discovery linking feature. When the participants had the discovery linking feature, they could click a single button, which looked for other words with matching definition keywords and synonyms. This meant that the participant could effectively do both of these actions with the click of one button when they had access to the linking features. This could likely explain why we saw decreased usage of both definition/synonym access and the number of searches when the linking feature was present. Lastly, our analysis of RQ 3.2 found that when clustering features were present, the participants used fewer of the linking features during the analysis process. This, of course, is due to the participant creating fewer links by using the clusters to create the grouping instead of using the links.

RQ 3.3: ***How do the new organizational features affect cognitive load while completing a sensemaking task?***

To address RQ3.3, we began by analyzing the NASA-TLX responses. We found no effect of *Clustering*, *Linking*, or *Clustering*Linking* on **Mental Demand**. We found an effect of *Linking* on Physical Demand ($F(1,90)= 5.94$, $p = 0.017$, $\eta_p^2 = 0.06$), where when *Linking* was present, we saw a higher **Physical Demand**. There were no effects of *Clustering* or *Clustering*Linking* on Physical Demand. We found an effect of *Clustering* on **Temporal Demand** ($F(1,93) = 6.74$, $p = 0.011$, $\eta_p^2 = 0.07$), where when *Clustering* was present, we saw lower Temporal Demand. There were no effects of *Linking* or *Clustering*Linking* on **Temporal Demand**. We found an effect of *Linking* on **Performance** ($F(1,90) = 4.46$, $p = 0.037$, $\eta_p^2 = 0.05$), where when *Linking* was present, we saw higher responses (closer to failure), on perceived performance. There were close to significant effects on *Clustering* ($F(1,90) = 3.90$, $p = 0.051$, $\eta_p^2 = 0.04$), with higher responses (closer to failure), on perceived performance. There was no effect of *Clustering*Linking* on Performance. We found no effect of *Clustering*, *Linking*, or *Clustering*Linking* on **Effort** or **Frustration**. Box plots visualizing the results of the NASA TLX can be seen in figures 6.20 and 6.21.

In addition to looking at the NASA-TLX scores, we also asked the participants two other questions to gauge overall cognitive load across the conditions. These two questions were part of the post-task questionnaire and scaled from Strongly Disagree to Strongly Agree. **This task was easy to complete -** Our analysis revealed an effect of Linking ($F(1,90) = 9.98$, $p = 0.002$, $\eta_p^2 = 0.10$), where when *Linking* was present, we saw lower scores in response to this question (i.e., more disagreement). The second question was **The features provided in this condition supported my sensemaking in completing the task**. Our analysis revealed an effect of *Clustering* ($F(1,93) = 18.04$, $p < 0.001$, $\eta_p^2 = 0.16$), where when *Clustering* was present, we saw higher responses to this question (i.e., more agreement). There were no effects of *Linking* or *Clustering*Linking* on question response. Box plots visualizing the results from the special questions on the post-task questionnaire can be seen in figure 6.22.

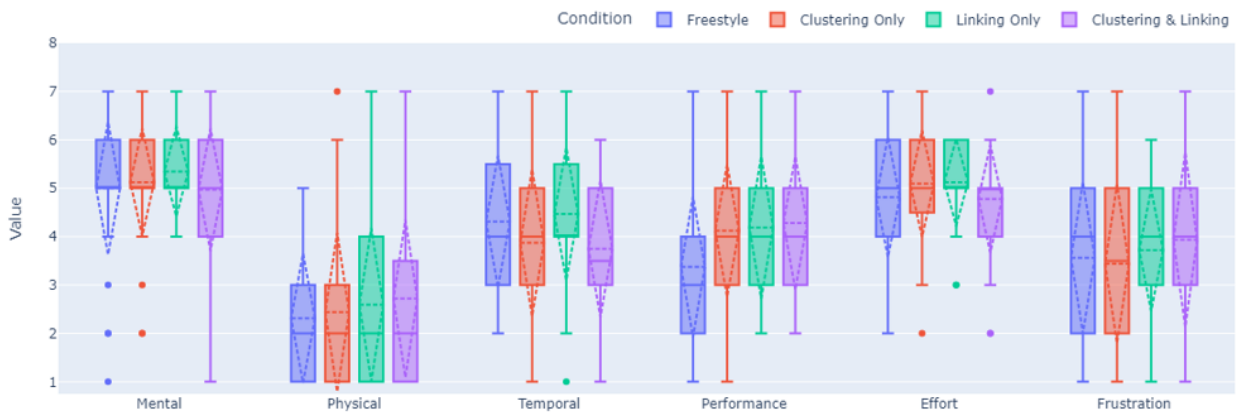
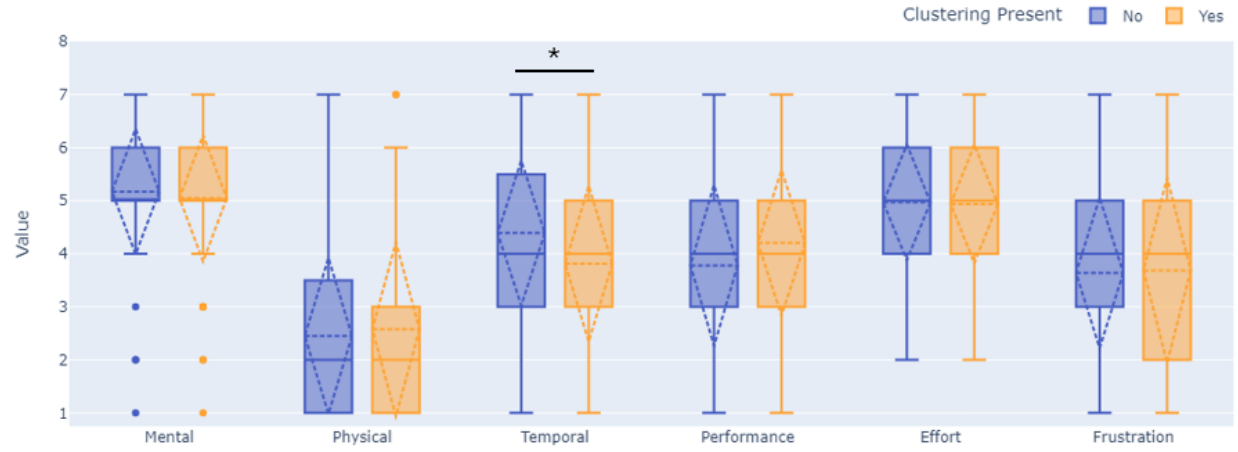


Figure 6.20: NASA TLX Result Box plots broken down by Condition. The dotted line and diamond visualize the mean and Standard Deviation. Higher scores represent a higher workload. Whiskers show the spread of the data without outliers.

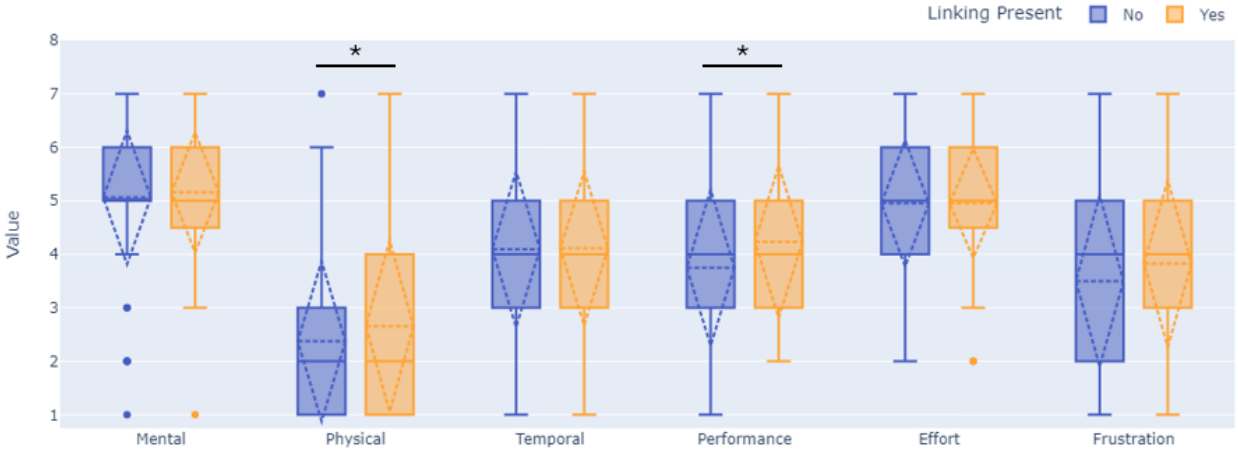
RQ 3.3 Discussion

In RQ 3.3, we asked how the new organizational features affected the cognitive load while completing a sensemaking task. Overall, our analysis revealed no differences across experimental conditions in the NASA-TLX responses. When looking at our two-by-two variables of the presence of clustering/linking, we identified that participants felt less rushed (temporal demand) when clustering was present. Additionally, we found that participants felt the task was more physically demanding when linking was present, and the participants felt they performed worse on the task when linking was present. In addition to the NASA-TLX data for mental task load, we also asked some custom Likert scale questions, which, again, revealed no difference across experimental conditions. However, we again identified some significant trends when looking at our two-by-two variables where participants felt the features provided were more beneficial to their sensemaking when Clustering was present. Moreover, when Linking was present, participants felt the task was harder to complete.

A few interesting factors could explain these trends we identified. One potential reason why the participants felt less temporal demand when clustering was present could have been that the clustering feature provided a way of bulk moving documents within the space, which may have provided a perceived efficiency boost when moving and reorganizing information in space while completing the experimental task. Additionally, participants reported higher physical demand when we look at the linking features. This could be due to the linking feature taking more interactions to use than clustering features or no linking features. For the users to create a link between two documents, they must select the anchor point from one document, then move the ray to the anchor point of the next document and select. Additionally, anchor points and links are harder to select within IST due to their size. Meanwhile, in the clustering (and freestyle) conditions, the participant only needs to move



(a) Clustering Present on NASA TLX.

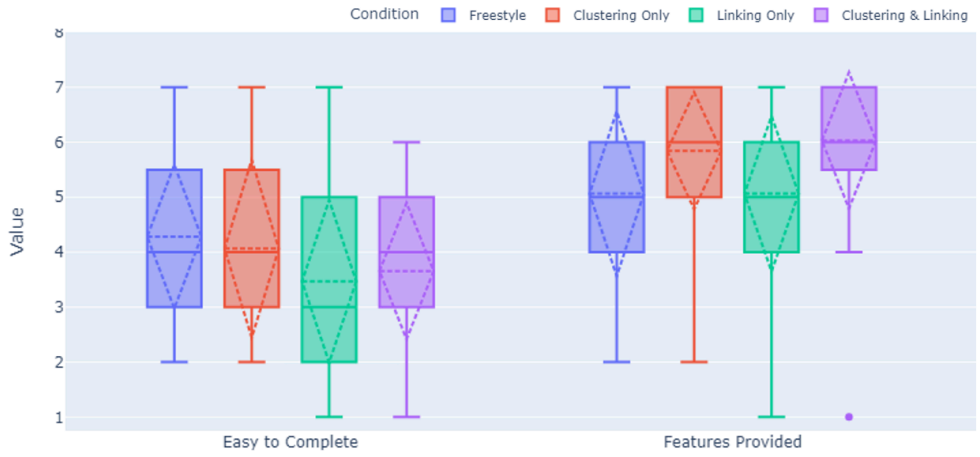


(b) Linking Presence NASA TLX.

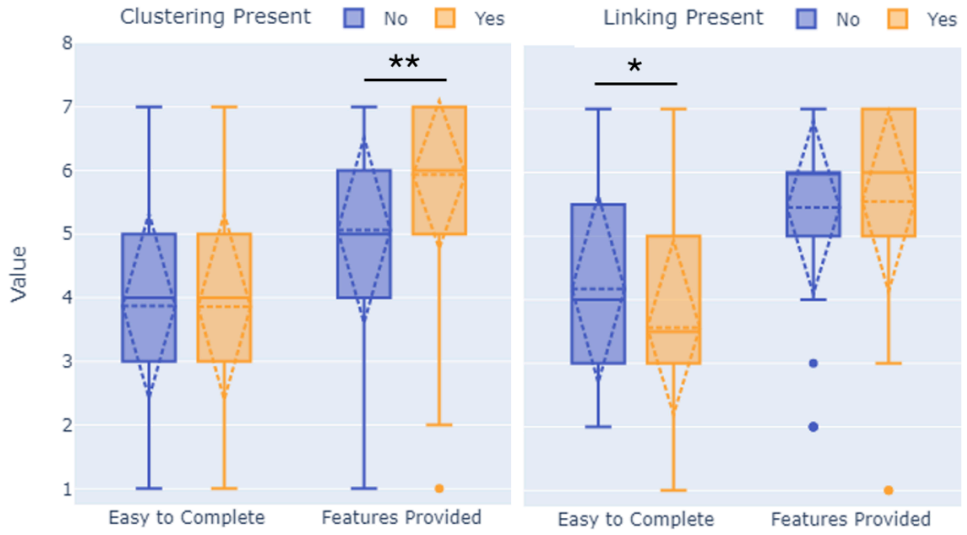
Figure 6.21: NASA TLX Result Box plots broken down by Presence of Clustering or Linking. Higher scores represent a higher workload. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05

the document and release the trigger. This could have also led to the difference we detected in the custom Likert scale questionnaire response *Easy to Complete*.

Overall, for cognitive load, few significant differences were identified between the responses and when Clusters and Linking were present. In addition to this, the differences identified were small; for instance, averages were still within 1 Likert scale point of each other. Additionally, the range of the responses we received related to mental task load was large as well. This could likely be due to the experimental task being challenging to complete, as suggested by the high (closer to 7) responses on Mental Demand, Performance, and Effort on the NASA-TLX questionnaire.



(a) Special Questions Condition Results Breakdown



(b) Special Questions Clustering and Linking Presence Results Breakdown

Figure 6.22: Results from special questions on the post-task questionnaire. Top - Condition specific breakdown, bottom - Clustering and Linking Presence breakdown. The scale is Strongly Disagree (Lower Scores) and Strongly Agree (Higher Scores). Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05 and ** ≤ 0.001

RQ 3.4: ***How do the new organizational features affect the system's overall user experience?***

The UEQ Questionnaire [55] contains 26 questions, which combine into 6 categories. Box plots visualizing the results of the UEQ can be seen in figures 6.23 and 6.24.



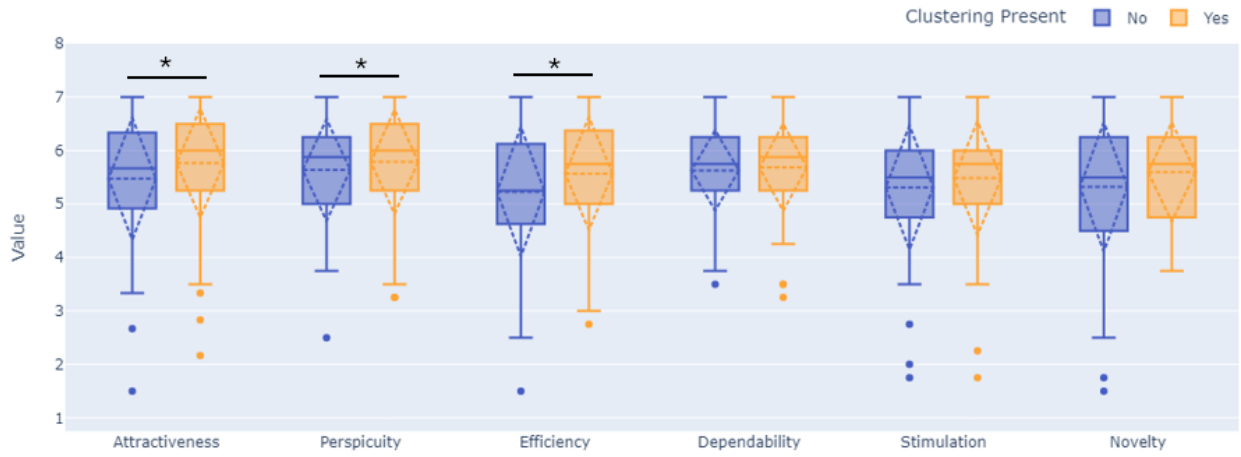
Figure 6.23: UEQ Results Box plots broken down by Condition. The dotted line and diamond visualize the mean and Standard Deviation. Higher scores indicate better usability. Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05

Attractiveness is the measure for the overall impression of the product and if the users like it or dislike it. Our analysis indicated a significant effect of *Clustering* ($F(1,93) = 7.74$, $p = 0.007$, $\eta_p^2 = 0.08$), and *Clustering*Linking* ($F(1,93) = 4.02$, $p = 0.048$, $\eta_p^2 = 0.04$) on *Attractiveness*. Post hoc analysis revealed that when clustering was present, attractiveness scores were higher ($p = 0.007$). Additionally, *Freestyle* scored higher than *Linking Only* ($p = 0.049$), *Clustering Only* scored higher than *Linking Only* ($p = 0.015$), and *Clustering & Linking* scored higher than *Linking Only* ($p = 0.006$) on *Attractiveness*.

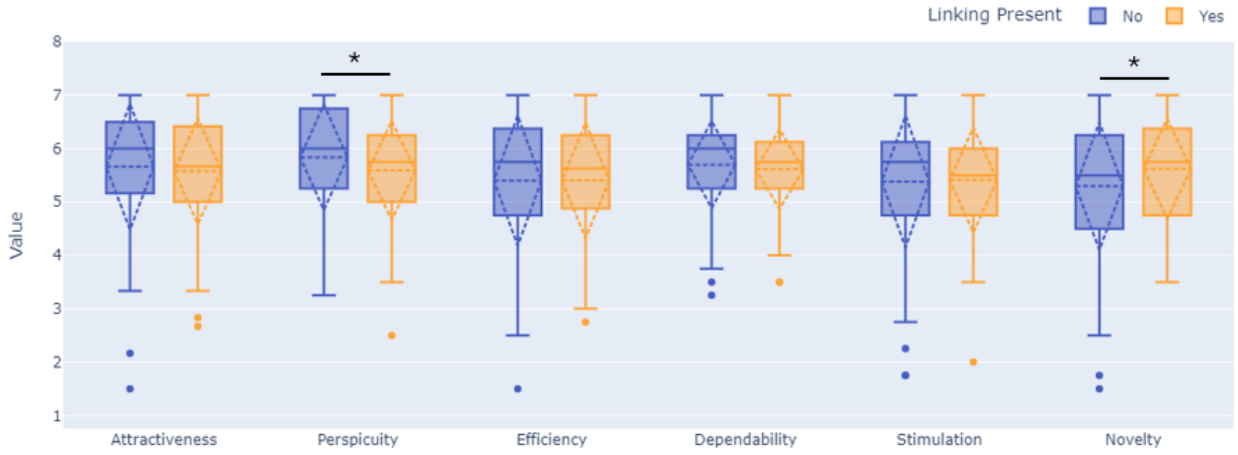
Perspicuity is the measure of how easy it is to learn how to use the system. The analysis indicated a significant effect of the presence of *Clustering* ($F(1,90) = 4.30$, $p=0.041$, $\eta_p^2 = 0.05$), presence of *Linking* ($F(1,90) = 9.80$, $p = 0.002$, $\eta_p^2 = 0.10$), and *Clustering*Linking* ($F(1,90)= 7.82$, $p = 0.006$, $\eta_p^2 = 0.08$) on Perspicuity. Post hoc analysis revealed that when clustering was present, perspicuity scores were higher ($p=0.04$), and when linking was present, perspicuity scores were higher ($p=0.002$). Additionally, *Freestyle* scored higher than *Linking Only* ($p=0.0004$), *Clustering Only* scored higher than *Linking Only* ($p = 0.004$), and *Clustering & Linking* scored higher than *Linking Only* ($p = 0.006$) on Perspicuity.

Our analysis revealed a significant effect of the presence of *Clustering* ($F(1,90) = 7.84$, $p = 0.006$, $\eta_p^2 = 0.08$) on **Efficiency**. Where when clustering was present, Efficiency scores were higher ($p = 0.006$). There were no effects of the presence of *Linking* or *Clustering*Linking* on Efficiency. Our analysis identified no effects of the presence of *Clustering*, *Linking*, or *Clustering*Linking* on **Dependability** or **Stimulation** scores. Lastly, our analysis indicated a significant effort of the presence of *Linking* ($F(1, 93) = 4.68$, $p = 0.033$, $\eta_p^2 = 0.05$) for **Novelty**, where when Linking was present we received higher Novelty Scores ($p = 0.033$).

Qualitative Feedback



(a) Clustering Presence UEQ Results Breakdown



(b) Linking Presence UEQ Results Breakdown

Figure 6.24: UEQ Results Box plots broken down by Present of Clustering or Linking. Higher scores indicate better usability. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05

In the post-study interview, we ask the participants a series of questions on their experiences during the study. One question we asked was **What feature was most effective in assisting you in finding the connections?** 17 participants (53.13%) stated that the discovery linking feature was most helpful, seven participants (21.88%) stated that clustering was the most helpful, six participants (18.75%) stated that the definitions & synonyms were the most helpful, one participant stated the ability to organize information around them (freestyle), and one participant stated the ability to search. We also asked the participants **Which feature/s were most effective in helping you stay organized during your analysis?** 28 participants (87.50%) stated that the clustering features were most helpful. Two participants stated that the freestyle (ability to organize the information in 3D space around them) was most helpful, and two participants stated that Clustering & Linking together were the most effective. P16 responded that *It is kind of hard to pick one. The links are really useful in organizing your thoughts on which connections might belong together, but the clustering was helpful for physically organizing the information.*

We also asked each participant **Which condition best supported your strategy in solving the connections task?** 20 participants stated *Clustering & Linking*, seven participants stated *Clustering Only*, three participants stated *Freestyle*, and two stated *Linking Only*. Participants who favored *Freestyle* stated that *I think my favorite was just free mode, it made it easy to move items and removed a layer of complexity - P23*, and *It allowed me to just focus on the task - P19*. Participants who liked *Clustering Only* suggested that *Linking was too busy. So I think I would pick just Clustering - P32.* The participants who liked *Linking Only* the most felt it *was easier to focus on the connections - P4.* Overall, the sentiment of those who preferred *Cluster & Linking* together was that *You can't have them alone, I think they are the best together - P6.*

We also asked questions targeted towards getting feedback on each of the features in general, **What were your thoughts on the _____ (Clustering, Manual Linking, and Discovery Linking) features and how they currently support your sensemaking?**

Clustering - We received very positive feedback overall on the clustering features. Participants felt it was *Very simple and easy to understand. (P11)* and helped with overall efficiency because they were able to *able to move multiple blocks at once (P10).* Additionally, many of the participants felt that the clustering features were *more visually appealing (P12)* than the linking features. The general sentiment from the participants was that the clustering features were helpful for *Organizing your solution (P17).* However, there were some concerns related to clustering. For instance, P6 stated that *Don't think it [clustering] did well in terms of supporting my sensemaking, but I guess the quality of life, it was very convenient to have.* Additionally, P16 stated that sometimes they struggled to *Remember that they were there* when the clusters were outside of their field-of-view.

Linking - The feedback we received on the linking feature was mixed. Participants who praised the linking features stated that it was overall more helpful for analysis, with P4 stating that linking was *Helpful for analyzing, but not as much for organizing (P4)*, and P6

stating that *Linking was much more relevant to the strategy.* Additionally, the participants who praised the linking feature expressed that it was helpful in showing uncertainty in analysis, with P5 stating it was *helpful if I was unsure about something and I wanted to connect two things* and P14 stating that linking *gave a card a purpose to be in the place that I put it in. So it wasn't just like a floating card in space.* Some participants shared a *middle of the road* sentiment with the linking features stating things like there were *No issues about it. I didn't really see the purpose of it. Or maybe it's just the way that my brain thinks... So, I didn't find it particularly useful, but I saw the potential for it (P11).* On the other hand, participants who criticized linking overall felt that it was *it's tedious to do the lines between each thing (P12)*, felt that the *lines everywhere kind of over-stimulated my brain a little bit (P19)*, and that *if it is just linking alone, things tend to get messy (P6).*

Discovery Linking - The feedback we received on the discovery linking features was positive. Overall, the participants felt that the discovery tool *was really useful for analysis and seeing stuff that was associated with words (P4)* and *Very useful as a tool that made up for a lack of understanding (P16).* The participant who shared similar feelings toward the discovery feature felt that it was helpful in *reducing your data set (P22)* and *opened my eyes to different possibilities (P14).* However, there were some concerns around the discovery linking tool with P21 stating that *I think it's very, very useful, but I also, I feel like I spent too much time using just that.* Additionally, P20 expressed concerns that sometimes *the connections it made were really weird*, and P6 stated that *everything gets too clustered around the diagram. I am so glad that there is a button to put everything back where it was.*

RQ 3.4 Discussion

RQ 3.4 asked how the new organizational features affected the participant's user experience with the IST system. For this question, we analyzed the UEQ and the interviews. Our analysis of the UEQ responses revealed that overall, *Linking Only* was the least attractive and hardest to learn (Perspicuity) of all of the conditions. Additionally, the analysis found that when linking was present, the participants ranked the overall novelty higher than when linking was not present. Additionally, in our two-by-two analysis, we found that when clustering was present, the participants rated the overall attractiveness, perspicuity, and efficiency higher than when clustering was not present.

One possible reason why, when clustering was present, we saw higher responses to attractiveness and efficiency could be explained within the participant's interview responses. A general sentiment from the interviews is that the clustering features were helpful for organizing information during the analysis. Additionally, the participants perceived the bulk move capabilities of clustering as very helpful in moving the information around the space, which could have led to higher efficiency.

As for the results related to the presence of linking, one possible explanation could be that since the linking feature was more physically demanding, this also made it harder to

learn. As previously stated, creating a link required more interactions than the freestyle/-clustering features. Additionally, the linking conditions included the discovery of linking features, which also may have caused the linking to be harder to learn how to use than the freestyle/clustering-only conditions. However, in contrast, the discovery linking feature could have also been the reason why, when linking was present, the results were higher in the novelty section. Another possible reason for the lower score on perspicuity could be that the participants were unsure of how to utilize the linking features during their analysis. In the interviews, some of the responses to the linking features highlighted that the lines were over-stimulating and messy or that they couldn't see a use case for the linking feature. Overall, challenges in understanding how to use the feature to benefit their analysis process could have led to lower scores on perspicuity.

One interesting aspect of the UEQ responses is that the overall median and averages were relatively high across all 6 combined UEQ factors. In our UEQ, we used a 7-point scale, meaning that the neutral response was a 4. In all of our analyses, the responses had median and mean values above 5. Overall, this is a good indicator that the overall usability of all of the conditions was high. This could have also led to the smaller effects that we saw across the UEQ responses.

In the interviews, we found that overall the participants felt that the clustering features (87.50%) were most helpful for organization, and the discovery linking (53.13%) was most helpful in helping the participants find the hidden connections. The general sentiment of the cluster features was very positive, and the participants liked that they could bulk move the documents. The linking features were a bit more divisive. Overall, the participants shared in the interviews that they were helpful for analyzing the dataset but were also *tedious*, *messy*, and *over-stimulation*.

RQ 3.5: *How do the new organizational features affect task performance?*

We looked at multiple measures to understand how the new features affected the task performance.

First, in the post-task questionnaire, we asked a question targeted towards gauging the participant's perceived performance on the task scaled from Strongly Disagree to Strongly Agree (7-Point Likert), **I feel confident in my solution to this task.** Our analysis revealed that there was an effect of the presence of *Linking* on the response variable ($F(1,93) = 4.35$, $p = 0.039$, $\eta_p^2 = 0.04$), where when Linking was present, we saw lower scores in participant confidence in their solution, as seen in figure 6.25.

Next, we examined the participants' overall performance on the tasks. To measure performance, we graded the solutions submitted by the participants based on the ground truth of the original connections game from which the tasks were created. The grading was based on two different aspects of the solutions: the number of documents correctly grouped and

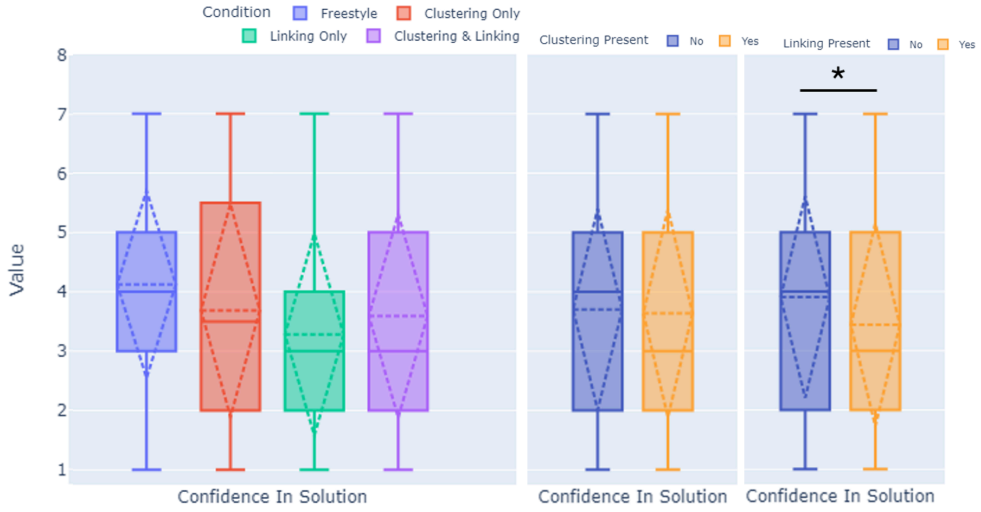


Figure 6.25: Results from special questions on confidence in solution on post-task questionnaire. The scale is Strongly Disagree (Lower Scores) and Strongly Agree (Higher Scores). Whiskers show the spread of the data without outliers, and significance is denoted with * $\leq$ 0.05.

the correctness of the labels generated by the participants. For scoring the document groupings, we used the grading scheme given in table 6.4. To grade the labels generated by the participant, we evaluated the labels based on the ground truth of the connections game. For example, the label for the grouping of documents *Brush, Canvas, Easel, and Palette* was *Painting Accessories* in the connections game solution. In the grading, we evaluated the participant labels based on similar sentiment, so for this grouping of documents, some examples of accepted labels from the participants were *painting supplies*, *artist's tools*, or *things you paint with*. If the label sentiment did not match the ground truth, for example, *colors*, no point was awarded for the label. Additionally, for grading the labels, the participant needed to have at least 2 correctly grouped documents for the label to receive a point. For grading, we also gave participants a bonus of 1 point for a fully correct connection (4 correct documents with a correct label), as seen in table 6.4.

Document Breakdown	Points Awarded	Points Awarded for a Label with Correct Sentiment
4 Correct, 0 Wrong	4	**2**
3 Correct, 1 Wrong	3	1
2 Correct, 2 Correct (2 documents correct from two different connections)	3	1
2 Correct, 2 Wrong	2	1

Table 6.4: Modified Connections Task Scoring Rubric.

We found no effects of *Condition* ($F(3,84) = 0.40$, $p = 0.750$, $\eta_p^2 = 0.01$), *Experimental Task Dataset* ($F(3,84) = 0.53$, $p = 0.656$, $\eta_p^2 = 0.02$), or an interaction effect of *Condition*Dataset* ($F(9,85) = 0.58$, $p = 0.813$, $\eta_p^2 = 0.06$) on task performance. We also evaluated the presence of *Clustering* and *Linking* and found no effect of *Clustering* ($F(1,93) = 1.14$, $p = 0.288$, $\eta_p^2 <$

0.001) or *Linking* ($F(1,93) = 0.032$, $p = 0.858$, $\eta_p^2 < 0.001$). The results of task performance can be seen in figure 6.26. Another factor we looked at was the total number of fully correct groups (4 correctly organized documents) formed by the participants in their analysis. We found no effect of *Condition* ($F(3,84) = 0.15271$, $p = 0.92768$, $\eta_p^2 < 0.001$), *Dataset* ($F(3,84) = 0.72$, $p = 0.545$, $\eta_p^2 = 0.02$), or *Condition*Dataset*($F(9,85) = 0.45$, $p = 0.905$, $\eta_p^2 = 0.05$) on the number of fully correct groups identified. We also evaluated the presence of *Clustering* and the presence of *Linking* on the number of fully correct groups and found no effect of *Clustering* ($F(1,93) = 0.22$, $p = 0.636$, $\eta_p^2 < 0.01$), or *Linking* ($F(1,93) = 0.38$, $p = 0.540$, $\eta_p^2 < 0.01$). The results of the number of fully correct groups can be seen in figure 6.27.

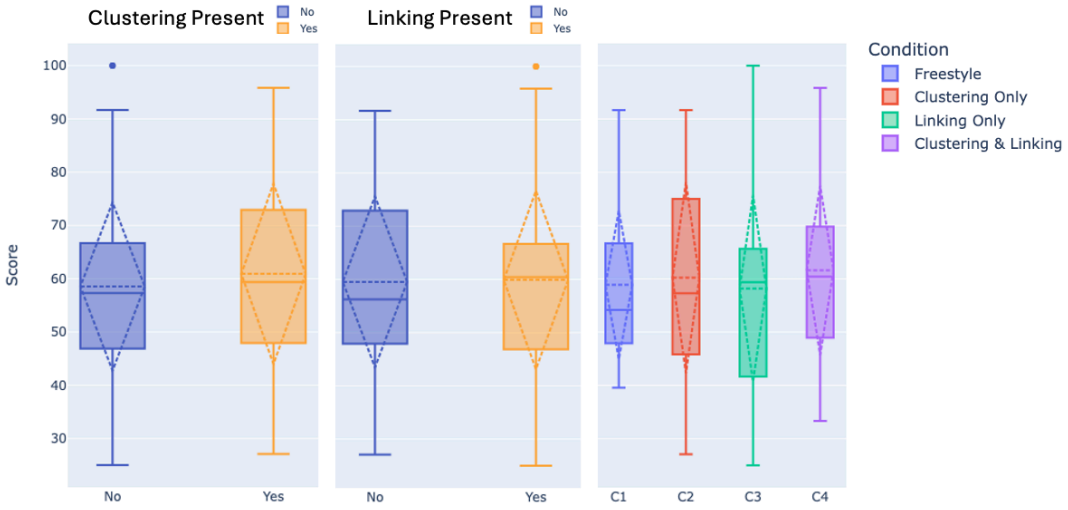


Figure 6.26: Task Performance Breakdown for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers. 100 represents a perfect score on the task.

Lastly, we also evaluated the total time taken to submit a solution to the task. The task was designed to take about 15 minutes to complete, and then the participants had a buffer of about 5 minutes to get their solution submitted. We identified an effect of *Experimental Condition* ($F(3,84) = 6.35$, $p < 0.001$, $\eta_p^2 = 0.18$) on time to submit the solution. There was no effect of *Experimental Task Dataset* ($F(3,84) = 1.99$, $p = 0.121$, $\eta_p^2 = 0.07$) or *Condition*Dataset* ($F(9,85) = 1.79$, $p = 0.082$, $\eta_p^2 = 0.16$). Contrasts revealed that *Freestyle* took less time to submit a solution than *Linking Only* ($p = 0.0004$), and *Freestyle* took less time to submit a solution than *Clustering & Linking* ($p = 0.017$). We also found an effect when *Linking* is present ($F(1,93) = 16.61$, $p < 0.001$, η_p^2 0.15), showing that when linking is present, it took longer to submit a solution. The results of the time taken to submit the task can be seen in figure 6.28.

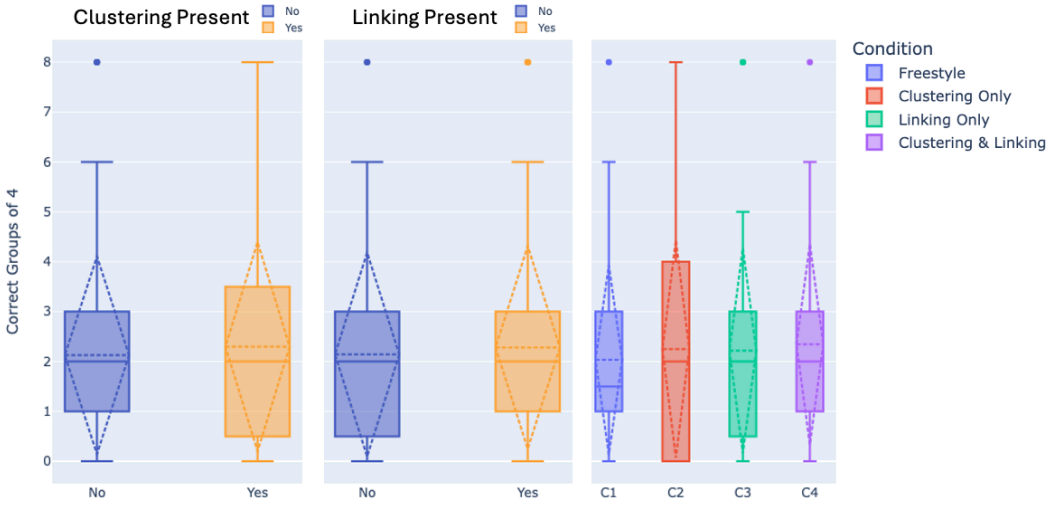


Figure 6.27: Count of correct groups of four breakdowns for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers.

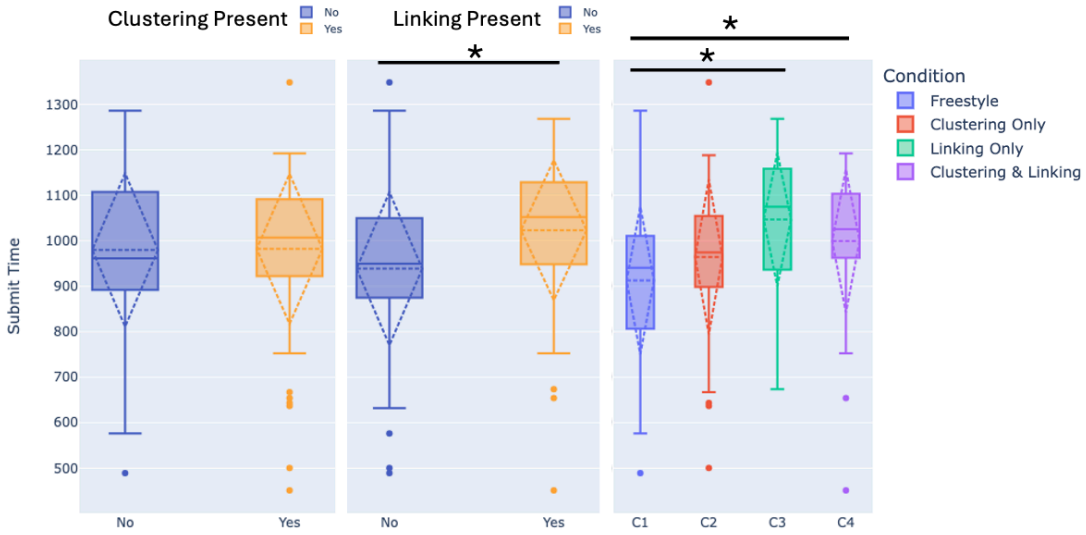


Figure 6.28: The time taken to submit the solution for the presence of clustering, presence of linking, and condition-specific breakdown. The dotted line and diamond visualize the mean and Standard Deviation. Whiskers show the spread of the data without outliers, and significance is denoted with * ≤ 0.05

RQ 3.5 Discussion

In RQ 3.5, we asked how the new organizational features affected task performance. Our analysis revealed two significant effects related to task performance. First, the analysis

revealed that when linking was present, the participants felt less confident in their solutions to the task. Second, the analysis revealed significant differences in the time to submit the task, where when linking was present, the participants took longer to submit their solutions. Additionally, we identified experimental condition level significance with *Linking Only* and *Clustering & Linking* taking longer to complete than *Freestyle* .

One possible reason for the longer submission times when linking was present could be that linking involved more interaction steps compared to the clustering and freestyle features. However, this explanation does not fully account for why participants in *Clustering & Linking* also exhibited longer submission times than those in *Freestyle*, despite using significantly fewer links in *Clustering & Linking* than in *Linking Only*. As discussed in Section 6.4.7, another plausible reason for the increased time could be the use of the discovery linking feature. This feature may have provided participants with new ideas on how to connect different documents within the dataset. However, the suggestions from the discovery linking feature might have caused participants to take longer in their analysis, as they needed to sift through various possibilities to determine which suggested connections were relevant to the task.

Regarding the differences in overall confidence in their solutions, we speculate on a few possible reasons why the presence of linking may have decreased confidence. One possible explanation is that participants may have over-relied on the discovery linking feature, leading them to follow red herrings. For instance, opening a discovery graph on the word "WATER" would identify "LIQUID" as a related word. However, these two words belong to different connections (Utilities vs. States of Matter), potentially misleading participants in their analysis. Another explanation could be that the linking conditions, which required more physical effort, distracted participants from the task, resulting in a perception of their overall solutions as less effective.

To explore these possibilities further, we examined the correlation between the usage of the discovery linking feature and task performance, measured by correctness. We looked at the correlation between discovery linking usage (when linking was present) and participants' scores, as well as the total number of correct groups of four. Our analysis found no correlation between discovery linking usage or retained documents from this feature and overall scores or the number of correct groups formed. While this does not provide a complete picture of whether the discovery linking feature led participants astray, it indicates that its usage did not significantly impact their performance on the task.

Additionally, we investigated whether there was a correlation between the total number of links utilized during analysis (which may have increased physical demand) and task performance or the number of correct groups of four. Again, no correlation was found between the number of links used and participants' scores or the number of correct groups. This suggests that the difference in overall confidence may be linked to another aspect of linking.

One potential aspect that could have affected overall confidence is the design purpose of the

linking feature. One of the design benefits of the linking feature is that it supports representing many-to-one relationships within the immersive space. Moreover, participants reported using the linking feature to represent uncertainty in a connection during interviews. It is possible that the ability to visually represent potential connections led to lower confidence in their solutions when the linking feature was present compared to when it was absent.

6.4.8 Key Takeaways

Our analysis revealed some interesting differences between the experimental conditions. In RQ 3.1, we found that *Linking Only* had the least tidy workspace and that the clustering feature led to greater total distances between the documents within the analysis space. RQ 3.2 revealed that when discovery linking was present, participants used it to replace definition/synonym lookup and execute manual searches. Additionally, the analysis of RQ 3.2 revealed that the linking features in *Linking Only* were not used as we expected. In RQ 3.3, we found that the clustering features helped lower the temporal demand, while the linking features increased physical demand and led to worse perceived performance on the task. In RQ 3.4, we found that *Linking Only* was ranked lower on attractiveness and perspicuity than all other conditions. Lastly, in RQ 3.5, we found that when linking was present, the participants took longer to submit their solutions.

Overall, throughout our analysis, when we identified significance, the differences were often very small (e.g., within 1 Likert scale point). One possible reason for these small differences could be related to the experimental task itself. We aimed to design a challenging sensemaking task comparable to previous studies in terms of the number of artifacts, and it was able to be completed in about 15 minutes. However, the overall performance on the task, with averages and medians around 60% correctness, indicated that it was challenging, as supported by the high scores on mental demand in the NASA-TLX. This high level of difficulty may have led participants to focus primarily on completing the task, resulting in them using the new experimental features in the most straightforward way rather than exploring their full potential.

Another takeaway from our analysis and the interviews suggests that linking features within IST needs more work. With higher physical demand for use, being visually overwhelming to participants, and harder to learn how to use, future work surrounding linking could be fruitful. For example, future work on designing an easier way to create links or studying graph layout options could make the linking feature more robust and easier for participants to use.

Another intriguing and thought-provoking interpretation of our results is that the Freestyle IST condition provides users with many benefits in the Connections-based sensemaking task, such that including the additional organizational features only marginally benefited the sensemaking process within the context of this experiment. Freestyle IST already provides an expansive space that allows users to organize information in any way they prefer

during analysis, which enhances schema formation. Thus, these new organizational features may have been less impactful than they otherwise might have been, leading to the smaller effect sizes we identified for the experimental task. Additionally, with no detected difference in overall performance, it is possible that while these new features were useful to the participants, as suggested in the UEQ results, they did not significantly enhance their performance on the experimental task. Based on participant feedback, we believe these new features contribute positively to the sensemaking process within a system like IST. However, future work exploring other sensemaking tasks within IST with these new features could shed more light on the overall benefits and changes to the sensemaking process.

6.5 Limitations & Future Work

As with all studies, our work is not without limitations. First, the experimental task we designed was centered around grouping information into groups of 4. This means that the task had less freedom for organizing the information than our prior work in IST, which could have led to certain organizational feature preferences in this study. Further studies could be conducted to evaluate how these new organizational features affect the sensemaking process in more traditional text-based sensemaking tasks. Another limitation of our study was the combination of the discovery linking feature with manual linking since the discovery linking feature is not specifically an organization feature. Throughout our analysis, we looked at the usage of linking features with the combined and separate usage of the manual links and discovery features. However, evaluating how linking alone affects sensemaking was not included as a condition in this study. Another possible direction for future work could be enhancing the linking features within IST. In this work, we developed a preliminary implementation of the linking feature; however, we did not apply any smart layouts to the links, which could help prevent visual clutter within the system. Evaluating different techniques for link management could be a fruitful next step. Another limitation of our work is surrounding the discovery linking features and the potential to cause confirmation bias within the analysis process. For instance, if the participant believes that one of the hidden connections is *Things with Teeth*, and they have already grouped the words *SAW*, *GEAR*, and *ZIPPER*, they might use the discovery link and find the word *CANINE*, which does not belong in the *Things with Teeth* connection, instead the last item is *COMB*. Future work exploring how discovery linking can be implemented in a way to help analysts identify relevant information while also prompting analysts to disprove wrong hypotheses [43] would be a valuable direction for future work.

Another limitation of this work was the experimental task length. With tasks around 15 minutes in length, the participants had less time to explore the usage of the new features and the system. A future study could be conducted to explore how the feature usage changes as the participants become more experienced in their usage. Another interesting direction for future work could be to consider the scalability of these new organizational features

to larger analysis datasets. In our study, the participants had 32 artifacts, which can be easily organized within the space around them without overlap issues. Exploring how these features are utilized with a larger dataset could explore the scalability of these features and their usage within IST.

6.6 Conclusions

This chapter explored how new organizational features affected sensemaking in an immersive analytics system. We explored five research questions that focused on spatial structures, strategies, mental task load, user experience, and task performance. We found that the *Linking Only* condition had the least tidy workspaces and that clustering features led to greater total distances between the documents within the analysis space. We identified changes in strategies between the experimental conditions, with different usage of features within the system. For mental task load, we found that clustering features helped lower the temporal demand, while the linking features increased physical demand and led to worse perceived performance on the task. For the systems user experience, we received high rankings overall, indicating a good user experience. However, *Linking Only* was ranked lower on attractiveness and perspicuity than all other conditions. Lastly, in task performance, we found no effect of the new organizational features on experimental task performance and found that when linking was present, the participants took longer to submit their solutions to the tasks. Using these findings, we developed an understanding of how these new organizational features affected the sensemaking process.

Chapter 7

Conclusions & Future Work

7.1 Conclusions

This dissertation focused on three main research questions aimed at exploring the utilization of immersive analytic systems in text-based sensemaking tasks and designing supplementary features to enhance analysis within these systems. This chapter examines each research question and presents the findings from the user studies conducted to address these questions.

7.1.1 RQ 1: How do sensemaking strategies evolve during multi-session analysis in Immersive Space to Think?

Previous work on understanding sensemaking within immersive analytics systems has largely focused on single-session sensemaking tasks. These research studies have provided many preliminary insights into how these systems can be used for analysis tasks. However, these single-session user studies do not evaluate or address how these systems can be used during multi-session analysis tasks where the participant has time to evolve their approach over the analysis task. To address research question 1, we conducted two exploratory user studies involving different participant populations. Our user studies focused on understanding how sensemaking evolves during multiple analysis sessions in IST. In these multi-session studies, we identified lower-level organization schemas used during the analysis process, including timelines, trash piles, network graphs, and scratch space usage. Additionally, we identified patterns in how the immersive spaces changed during the analysis sessions. As the participants got deeper into the sensemaking process, the overall workspace tidiness (using silhouette and CH scores) increased. We also identified participant movement patterns, with participants' overall movement decreasing into the late stages of the analysis process. However, while the overall movement decreased, we found that the participants still physically navigated around the space to reference information as needed while building the report of their findings in the *presentation* stage of the sensemaking process. We also found that there were two groups of participants, **Movers** and **Stationary** when it came to physically navigating the immersive space during the analysis process. Additionally, we compared *Novice* and *Professional* analysts to understand how these two populations compare in their sensemaking within IST. Overall, the findings from our comparison suggested that while there were interesting differences between the two populations, novice user populations are

able to provide novel insights into immersive analytics sensemaking in traditional user study settings.

7.1.2 RQ 2: How does familiarity affect initial strategies and layouts used for immersive sensemaking?

After understanding how the sensemaking strategies evolved during multi-session analysis within IST, we wanted to understand how familiarity with the analytic tool affected the strategies and layouts used for sensemaking. Specifically, we wanted to understand if the participants' initial spatial structure and strategies changed as they became more familiar with the tool. We performed an exploratory user study in which the participants completed three separate text-based sensemaking tasks during three separate sessions within IST. Our analysis revealed both high and low-level changes in spatial structures (schemas) and strategy across the sensemaking sessions. For instance, we found that 57.1% of participants used at least one different high-level spatial structure during their analysis and found that the participants utilized more of the full 360-degree space for document layout as they became more familiar with the IST system. We also found that the immersive spaces became more tidy as the participants became more familiar with the system. Additionally, 71.4% of participants reported using a different strategy for analysis between sessions one and two, and 35.7% between sessions two and three. Using our findings from this study, we presented a set of best practices for analysis within IST, which guided our feature design within our last research question, RQ3.

7.1.3 RQ 3: How can we support best practices for immersive sensemaking with additional organizational features?

To address RQ3, we developed new features to aid in organizing information while completing analysis tasks in IST. Using these features, we ran an expert feedback session to gather insights on the new features and how we could modify them to better support the participant's sensemaking. After modifying these new features, we conducted a user study to understand how these new features affected the overall sensemaking process within IST. In this study, we compared four conditions of IST, Freestyle, Clustering Only, Linking Only, and Clustering & Linking, to understand how the addition of these features to IST affected the spatial structures, strategies for analysis, cognitive load, user experience, and task performance. This study revealed that the Linking Only condition had the least tidy workspace and that the presence of the clustering feature led to greater total distances between the documents within the analysis space (i.e., spread of the documents). We also found that the clustering features helped lower the temporal demand during the experimental task, while the linking features led to increased physical demand and worse perceived performance on the task. We also found no effect of the new organizational features on the experimental task

performance when looking at the score of the experimental tasks. Additionally, our study results suggested the overall power of the baseline IST (i.e., freestyle) condition and its ability to support sensemaking within the context of the experimental task while emphasizing the importance of future work around the linking features to improve their overall usability.

7.2 Future Work Directions

7.2.1 Scalability

One unexplored area of research related to IST is scalability. Our work has primarily focused on small datasets containing 40 or fewer documents. However, real-world scenarios often involve datasets with hundreds, thousands, or even millions of items. How can IST be scaled to handle these larger document sets?

One approach could be to develop methods for minimizing information that is not currently being used for analysis, thus reducing cognitive load and visual clutter. Another idea is to implement multiple rooms or portals that users can enter and exit during the analysis, each containing different analysis themes. However, it remains unclear how having multiple analysis spaces within IST would affect spatial memory.

Another future direction could involve exploring the amount of immersive space available. Our current work has utilized finite amounts of immersive space, with room-scale tracking and physical navigation. However, immersive space can be theoretically infinite. In an infinite immersive space, virtual navigation methods such as teleportation could be used to access different areas. It would be interesting to investigate how much space users utilize for analysis, how spatial memory is affected, and how users partition the infinite space for different tasks.

Overall, there are numerous intriguing questions to address regarding the scalability of the IST concept. Exploring these questions could significantly enhance the utility and effectiveness of IST for large-scale data analysis.

7.2.2 Longitudinal Studies

Another unexplored aspect of IST is understanding how the usage of an immersive analytics system evolves in truly longitudinal studies. The work presented in this dissertation involved multi-session analyses, examining both single analysis tasks (Chapters 3 and 4) and multiple analysis tasks (Chapter 5). These studies provided insights into how immersive analytics systems support multi-session analysis and how participants' strategies and organization patterns changed as they became more familiar with the tool.

However, a key question remains: how would these systems be used as participants become even more *expert* with the system? In truly longitudinal studies, larger analysis tasks could be undertaken involving a greater number of documents, which would also address some aspects of system scalability.

Interesting questions in the context of longitudinal studies include examining trends and differences over time in the analysis process within the system. Additionally, longitudinal studies could explore memory recall in immersive analytics systems compared to traditional systems, providing insights into how these environments affect long-term information retention and usability.

Overall, conducting longitudinal studies with IST could reveal significant trends and patterns, contributing to a deeper understanding of how immersive analytics systems can support complex, long-term analytical tasks.

7.2.3 Collaboration

In this dissertation, we have focused on individuals completing analysis within IST. However, future work could address collaborative analysis within IST, building on the existing literature on Collaborative Immersive Analytics (CIA). There are many interesting questions and research directions related to collaboration within IST that warrant exploration.

One area to explore would be IST within the context of the time-space taxonomy, as outlined by Baecker [4], and the effects on collaborative sensemaking in IST. This taxonomy examines different aspects of time (synchronized vs. asynchronous) and space (co-located vs. distributed). Billinghurst et al. [8] highlighted the relevance of this taxonomy in CIA. For instance, how do co-located teams versus distributed teams utilize IST differently during a collaborative sensemaking task? Investigating these aspects could provide valuable insights into collaborative sensemaking in IST. Expanding this taxonomy, Sereno et al. [82] discussed role and technology symmetry, which looks at symmetric versus asymmetric roles (e.g., boss and employee) and technology (e.g., VR and AR). Understanding how these dynamics affect collaborative sensemaking can inform future designs of the IST concept, which could better support collaboration.

Additionally, territoriality, such as personal, shared or storage territories [81] of the analysis space during collaborative sensemaking tasks presents another interesting direction for future research. In IST, this could involve studying how virtual and physical spaces are divided and utilized by collaborators. Questions to explore include: How do teams negotiate and manage space within an immersive environment? What strategies do they employ to maintain organization and clarity in their shared workspaces?

Moreover, designing for individual-level control and group awareness, as suggested by Gutwin and Greenberg [40], is critical in collaborative workspace settings. Ensuring that individuals have the ability to manipulate and control their own data views and interactions while

maintaining an awareness of the group's activities can enhance both individual contributions and overall group performance. Research could focus on developing and evaluating interface features that balance personal views of the data with collective awareness.

7.2.4 Human-Machine Teaming

As we live in the world of big data, the amount of information we are presented when conducting analysis tasks is increasing. As we explore ways to scale immersive analytics systems like IST, incorporating recent advances in artificial intelligence (AI) and machine learning (ML), particularly large language models (LLMs), presents novel opportunities for enhancing human-machine teaming.

Integrating LLMs and other AI/ML technologies with immersive analytics systems could significantly augment IST's analytical capabilities, opening up new opportunities for human-machine collaboration. This fusion has the potential to streamline data processing, provide intelligent insights, and enhance the overall efficiency of the analysis process.

However, many open questions remain regarding human-machine teaming. For instance, how can we best leverage semantic interaction to guide AI algorithms in a way that aligns with human analytical processes (i.e., the sensemaking process)? Additionally, are there specific points within the sensemaking process for which human-machine teaming will provide better benefits in augmenting the human?

Another key aspect of building human-machine teaming is ensuring transparency and building trust in the system, which are critical. The users need to understand and trust the AI's contributions to feel confident in their collaborative assistant.

Some open questions for future research within human-machine teaming could be: How can we build transparency and trust in human-machine teaming within IST? What impact does the addition of human-machine teaming have on the outcomes of the analysis process? To what extent should analysts/machines be involved in the human-machine teaming process to maximize effectiveness and maintain control over the analysis?

Exploring these questions could lead to significant advancements in the design and implementation of immersive analytics systems, ultimately enhancing their ability to handle large-scale data and complex analysis tasks in a collaborative, transparent, and efficient manner.

Bibliography

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

[2] Christopher Andrews, Alex Endert, Beth Yost, and Chris North. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. *Information Visualization*, 10(4):341–355, 2011. doi: 10.1177/1473871611415997.

[3] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: Ordering points to identify the clustering structure. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, SIGMOD '99, page 49–60, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 1581130848. doi: 10.1145/304182.304187. URL https://doi.org/10.1145/304182.304187.

[4] Ronald M Baecker. *Readings in groupware and computer-supported cooperative work: Assisting human-human collaboration*. Morgan Kaufmann, 1993.

[5] Payel Bandyopadhyay. *Immersive Space to Think: the Role of 3D Immersive Space in Sensemaking of Textual Data*. PhD thesis, Virginia Tech, 2020.

[6] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):536–546, 2020. doi: 10.1109/TVCG.2019.2934803.

[7] Eric A Bier, Edward W Ishak, and Ed Chi. Entity workspace: an evidence file that aids memory, inference, and reading. In *International Conference on Intelligence and Security Informatics*, pages 466–472. Springer, 2006.

[8] Mark Billinghurst, Maxime Cordeil, Anastasia Bezerianos, and Todd Margolis. *Collaborative Immersive Analytics*, pages 221–257. Springer International Publishing, Cham, 2018. ISBN 978-3-030-01388-2. doi: 10.1007/978-3-030-01388-2_8. URL https://doi.org/10.1007/978-3-030-01388-2_8.

[9] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[10] D. A. Bowman and R. P. McMahan. Virtual reality: How much immersion is enough? *Computer*, 40(7):36–43, 2007. doi: 10.1109/MC.2007.257.

[11] Doug A. Bowman and Larry F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, page 35–ff., New York, NY, USA, 1997. Association for Computing Machinery. ISBN 0897918843. doi: 10.1145/253284.253301.

[12] Lauren Bradel, Chris North, Leanna House, and Scotland Leman. Multi-model semantic interaction for text analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 163–172, 2014. doi: 10.1109/VAST.2014.7042492.

[13] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[14] Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2015.

[15] Sarah Cohen, James T Hamilton, and Fred Turner. Computational journalism. *Communications of the ACM*, 54(10):66–71, 2011.

[16] Maxime Cordeil, Tim Dwyer, Karsten Klein, Bireswar Laha, Kim Marriott, and Bruce H Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE transactions on visualization and computer graphics*, 23(1): 441–450, 2016.

[17] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H Thomas, and Kim Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 71–83. ACM, 2017.

[18] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, page 135–142, New York, NY, USA, 1993. Association for Computing Machinery. ISBN 0897916018. doi: 10.1145/166117.166134. URL https://doi.org/10.1145/166117.166134.

[19] Kylie Davidson, Lee Lisle, Kirsten Whitley, Doug A. Bowman, and Chris North. Exploring the evolution of sensemaking strategies in immersive space to think. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–15, 2022. doi: 10.1109/TVCG.2022.3207357.

[20] Kylie Davidson, Lee Lisle, Ibrahim A. Tahmid, Kirsten Whitley, Chris North, and Doug A. Bowman. Uncovering best practices in immersive space to think. In *2023 IEEE*

International Symposium on Mixed and Augmented Reality (ISMAR), pages 1094–1103, 2023. doi: 10.1109/ISMAR59233.2023.00126.

[21] Maureen Duffy. Sensemaking in classroom conversations. *Openness in research: The tension between self and other*, pages 119–132, 1995.

[22] Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H Thomas. Immersive analytics: An introduction. *Immersive analytics*, pages 1–23, 2018.

[23] Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*, pages 754–768, 2021.

[24] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. Situated analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2015.

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

[26] Alex Endert, Russ Burtner, Nick Cramer, Ralph Perko, Shawn Hampton, and Kristin Cook. Typograph: Multiscale spatial exploration of text documents. In *2013 IEEE International Conference on Big Data*, pages 17–24. IEEE, 2013.

[27] Alex Endert, William Ribarsky, Cagatay Turkay, BL William Wong, Ian Nabney, I Díaz Blanco, and Fabrice Rossi. The state of the art in integrating machine learning into visual analytics. In *Computer Graphics Forum*, volume 36, pages 458–486. Wiley Online Library, 2017.

[28] Barrett Ens, Sarah Goodwin, Arnaud Prouzeau, Fraser Anderson, Florence Y Wang, Samuel Gratzl, Zac Lucarelli, Brendan Moyle, Jim Smiley, and Tim Dwyer. Uplift: A tangible and immersive tabletop system for casual collaborative visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1193–1203, 2020.

[29] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, et al. Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

[30] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[31] Alessandro Febretti, Arthur Nishimoto, Terrance Thigpen, Jonas Talandis, Lance Long, JD Pirtle, Tom Peterka, Alan Verlo, Maxine Brown, Dana Plepys, et al. Cave2: a hybrid reality environment for immersive simulation and information analysis. In *The Engineering Reality of Virtual Reality 2013*, volume 8649, pages 9–20. SPIE, 2013.

[32] Steven Feiner and Ari Shamash. Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, pages 9–17, 1991.

[33] Patrick Fiaux, Maoyuan Sun, Lauren Bradel, Chris North, Naren Ramakrishnan, and Alex Endert. Bixplorer: Visual analytics with biclusters. *Computer*, 46(8):90–94, 2013.

[34] Alexia Galati, Riley Schoppa, and Aidong Lu. Exploring the sensemaking process through interactions and fnirs in immersive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2714–2724, 2021. doi: 10.1109/TVCG.2021.3067693.

[35] Alexander Giovannelli, Lee Lisle, and Doug A. Bowman. Exploring the impact of visual information on intermittent typing in virtual reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 8–17, 2022. doi: 10.1109/ISMAR55827.2022.00014.

[36] Carsten Gorg, Zhicheng Liu, Neel Parekh, Kanupriya Singhal, and John Stasko. Visual analytics with jigsaw. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 201–202. IEEE, 2007.

[37] Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, and John Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2012.

[38] Carsten Görg, Zhicheng Liu, and John Stasko. Reflections on the evolution of the jigsaw visual analytics system. *Information Visualization*, 13(4):336–345, 2014.

[39] G. Grinstein, T. O'Connell, S. Laskowski, C. Plaisant, J. Scholtz, and M. Whiting. Vast 2006 contest - a tale of alderwood. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pages 215–216, 2006. doi: 10.1109/VAST.2006.261420.

[40] Carl Gutwin and Saul Greenberg. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98, page 207–216, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130090. doi: 10.1145/289444.289495. URL https://doi.org/10.1145/289444.289495.

[41] Markus Hällgren, Ola Lindberg, and Oscar Rantatalo. Sensemaking in detective work: The social nature of crime investigation. *International Journal of Police Science & Management*, 23(2):119–132, 2021.

[42] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[43] Richards J Heuer. *Psychology of intelligence analysis*. Center for the Study of Intelligence, 1999.

[44] James Hollan, Edwin Hutchins, and David Kirsh. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, 7(2):174–196, June 2000. ISSN 1073-0516. doi: 10.1145/353485.353487. URL https://doi.org/10.1145/353485.353487.

[45] Petra Isenberg, Danyel Fisher, Sharoda A Paul, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on visualization and Computer Graphics*, 18(5):689–702, 2011.

[46] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[47] Youn-ah Kang and John Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 21–30, 2011. doi: 10.1109/VAST.2011.6102438.

[48] Youn-ah Kang, Carsten Gorg, and John Stasko. Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pages 139–146. IEEE, 2009.

[49] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization*, pages 154–175. Springer, 2008.

[50] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann. *Mastering the information age solving problems with visual analytics*. Eurographics Association, 2010.

[51] Matthias Kipping, R Daniel Wadhwani, and Marcelo Bucheli. Analyzing and interpreting historical sources: A basic methodology. *Organizations in time: History, theory, methods*, pages 305–329, 2014.

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

[53] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. Virtual memory palaces: immersion aids recall. *Virtual reality*, 23:1–15, 2019.

[54] W. Kruger, C.-A. Bohn, B. Frohlich, H. Schuth, W. Strauss, and G. Wesche. The responsive workbench: a virtual work environment. *Computer*, 28(7):42–48, 1995. doi: 10.1109/2.391040.

[55] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*, pages 63–76. Springer, 2008.

[56] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer. Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2020. doi: 10.1109/TVCG.2020.3030450.

[57] Benjamin Lee, Maxime Cordeil, Arnaud Prouzeau, and Tim Dwyer. Fiesta: A free roaming collaborative immersive analytics system. In *Proceedings of the 2019 ACM international conference on interactive surfaces and spaces*, pages 335–338, 2019.

[58] Tianyi Li, Kurt Luther, and Chris North. Crowdia: Solving mysteries with crowdsourced sensemaking. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–29, 2018.

[59] Lee Lisle, Xiaoyu Chen, J.K. Edward Gitre, Chris North, and Doug A. Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 331–337, 2020. doi: 10.1109/VRW50115.2020.00073.

[60] Lee Lisle, Kylie Davidson, Edward J.K. Gitre, Chris North, and Doug A. Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537, 2021. doi: 10.1109/VR50410.2021.00077.

[61] Lee Lisle, Kylie Davidson, Edward J. K. Gitre, Chris North, and Doug A. Bowman. Different realities: a comparison of augmented and virtual reality for the sensemaking process. *Frontiers in Virtual Reality*, 4, 2023. ISSN 2673-4192. doi: 10.3389/frvir.2023.1177855. URL https://www.frontiersin.org/articles/10.3389/frvir.2023.1177855.

[62] Lee Lisle, Kylie Davidson, Leonardo Pavanatto, Ibrahim A. Tahmid, Chris North, and Doug A. Bowman. Spaces to think: A comparison of small, large, and immersive displays for the sensemaking process. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1084–1093, 2023. doi: 10.1109/ISMAR59233.2023.00125.

[63] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 588–597. IEEE, 2020.

[64] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. Effects of display layout on spatial memory for immersive environments. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):468–488, 2022.

[65] Weizhou Luo, Anke Lehmann, Yushan Yang, and Raimund Dachselt. Investigating document layout and placement strategies for collaborative sensemaking in augmented reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380959. doi: 10.1145/3411763.3451588.

[66] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. Where should we put it? layout and placement strategies of documents in augmented reality for collaborative sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3501946. URL https://doi.org/10.1145/3491102.3501946.

[67] J MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297. University of California Los Angeles LA USA, 1967.

[68] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. *Immersive Analytics*, volume 11190. Springer, 2018.

[69] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.

[70] Donald A Norman. Natural user interfaces are not natural. *interactions*, 17(3):6–10, 2010.

[71] Stanford School of Design, Date Accessed Jan 3rd, 2024. URL https://public-media.interaction-design.org/pdf/I-Like-I-Wish-What-If.pdf.

[72] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[73] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4. McLean, VA, USA, 2005.

[74] P. Reipschläger, T. Flemisch, and R. Dachselt. Personal augmented reality for information visualization on large interactive displays. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2020. doi: 10.1109/TVCG.2020.3030460.

[75] George Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten van Dantzich. Data mountain: using spatial memory for document management. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, page 153–162, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130341. doi: 10.1145/288392.288596.

[76] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[77] Alexander Rowden, Süleyman Aslan, Eric Krokos, Kirsten Whitley, and Amitabh Varshney. Waverider: Immersive visualization of indoor signal propagation. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pages 1–12, 2022.

[78] Alexander Rowden, Eric Krokos, Kirsten Whitley, and Amitabh Varshney. Exploring effective immersive approaches to visualizing wifi. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 732–740, 2023. doi: 10.1109/ISMAR59233.2023.00088.

[79] S. Sagiroglu and D. Sinanc. Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 42–47, 2013. doi: 10.1109/CTS.2013.6567202.

[80] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):1–20, 2020.

[81] Stacey D. Scott, M. Sheelagh T. Carpendale, and Kori Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, page 294–303, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138105. doi: 10.1145/1031607.1031655. URL https://doi.org/10.1145/1031607.1031655.

[82] M. Sereno, X. Wang, L. Besancon, M. J. Mcguffin, and T. Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2020. doi: 10.1109/TVCG.2020.3032761.

[83] Joseph Siegel. Did you take "good" notes?: On methods for evaluating student note-taking performance. *Journal of English for Academic Purposes*, 35:85–92, 2018.

[84] Richard Skarbez, Nicholas F Polys, J Todd Ogle, Chris North, and Doug A Bowman. Immersive analytics: Theory and research agenda. *Frontiers in Robotics and AI*, 6:82, 2019.

[85] Mel Slater and Sylvia Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 12 1997. doi: 10.1162/pres.1997.6.6.603. URL https://doi.org/10.1162/pres.1997.6.6.603.

[86] John Stasko, Carsten Gorg, Zhicheng Liu, and Kanupriya Singhal. Jigsaw: supporting investigative analysis through interactive visualization. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138. IEEE, 2007.

[87] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764, 1968.

[88] Ibrahim A Tahmid, Lee Lisle, Kylie Davidson, Chris North, and Doug A Bowman. Evaluating the benefits of explicit and semi-automated clusters for immersive sensemaking. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 479–488. IEEE, 2022.

[89] J.J. Thomas and K.A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006. doi: 10.1109/MCG.2006.5.

[90] Karl E Weick, Kathleen M Sutcliffe, and David Obstfeld. Organizing and the process of sensemaking. *Organization science*, 16(4):409–421, 2005.

[91] Sean Michael White. *Interaction and presentation techniques for situated visualization*. Columbia University, 2009.

[92] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. Embedded data representations. *IEEE transactions on visualization and computer graphics*, 23(1):461–470, 2016.

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

[94] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146, 2011.

[95] Hao Wu, Michael Mampaey, Nikolaj Tatti, Jilles Vreeken, M Shahriar Hossain, and Naren Ramakrishnan. Where do i start? algorithmic strategies to guide intelligence analysts. In *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics*, pages 1–8, 2012.

[96] Yalong Yang, Tim Dwyer, Kim Marriott, Bernhard Jenny, and Sarah Goodwin. Tilt map: Interactive transitions between choropleth map, prism map and bar chart in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4507–4519, 2021. doi: 10.1109/TVCG.2020.3004137.

[97] Robert B Zajonc. Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2p2):1, 1968.

[98] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE transactions on visualization and computer graphics*, 24(1):340–350, 2017.

Appendices

Appendix A

Exploring the Evolution of Sensemaking Strategies in Immersive Space to Think

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 #20-762)

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

Other Study Contact(s):

Kylie Davidson
Department of Computer Science, Virginia Tech
Email: kyliedavidson@vt.edu

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

Douglas A. 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 or corrected (glasses or contacts) vision with strong English skills and an interest in national security.

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 three 75 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 (90 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 (20 minutes, split into two separate sessions): During this phase we will perform a semi-structured interview on your experience in the daily study phase.

- Post-study Phase (20 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 three 75-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 an Intelligence Analysis training dataset 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 intelligence report data from a fictitious Intelligence Analysis training task. Using these sources, you will be asked to write both an outline of a paper and a report that addresses what action or actions the terrorists are planning, where they will occur, and when they will occur. During session one, you will be asked to simply analyze the dataset for the terrorist action or actions. During session two, you will continue your analysis and develop your report outline. Lastly, during session three you will use the outline you created to generate a report to address your findings. 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. After each session, we will conduct a semi-structured interview during the "Post-session" or "Post-study" phase. This interview will ask you about your experiences using the software. The "Post-study" 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 Risk)

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 fictional terrorist attacks, which 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 $75.00 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 - $25.00, 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.

Signature Block for Capable Adult

We are taking audio, screen, and video recordings of your participation in this research. Please initial the following statement to provide your permission for these recordings:

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

_____

A.2 Questionnaires & Interviews

A.2.1 Pre-Study Questionnaire

1. Gender (Make only one)

o Male

o Female

o Other

o Prefer not to share

2. Age

_____

3. What is your major of study?

_____

4. How are you feeling today?

_____

5. Do you wear glasses or contact lenses?

o Neither

o Glasses

o Contact Lenses

6. Handedness:

o Right-Handed

o Left-Handed

o Ambidextrous

7. Rate your fatigue level:

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

8. Rate your expertise with computers

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

9. How often do you use computers 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? (console or on the computer?)

o Not at all

o Once a month

o Once a week

o Several Times per Week

o Daily

12. How many times have you tried virtual or augmented reality?

o Never

o Once or Twice

o 3 - 10 Times

o More than 10 times

13. Comments:

_____

14. Rate your experience watching 3D stereo movies, in theaters and on TV:

1 2 3 4 5 6

Poor (Nausea, Headaches, etc.) o o o o o o Excellent

15. Comments:

_____

16. During a sensemaking task, which of the following do you use? (An example of a sensemaking task is reading documents and writing a research paper)

o Highlights

o Annotations

o Note Taking

o Other _____

17. Please explain your traditional approach to writing a research paper that analyzes a set of text documents.

_____

_____

_____

_____

A.2.2 Session 1 Interview

1. What is the overall strategy you used for analysis today?
2. Did your strategy for analysis change over the course of this session?
3. Describe to me the spatial layout you have formed thus far.
4. Did your spatial layout change over the course of this session?

A.2.3 Session 2 Interview

1. Did your strategy for analysis change over the course of this session?
2. Did the strategy you used today differ from the strategy you used during session one? If so, How?
3. Describe the spatial layout you have formed thus far?
4. Did the spatial layout you used today differ from the spatial layout you used during session one? If so, How?
5. Describe the transformation of your spatial layout to your report outline?

A.2.4 Session 3 Interview

1. Over the course of this analysis, how did you organize your documents to extract meaningful information?
2. Does your spatial layout reflect the organization of your report? If so, How?
3. Did your spatial layout change over the course of multiple sessions? If so, How?
4. Do you see advantages to using a system like IST over traditional analysis approaches for the sensemaking task?

A.3 Post-Study Semi-Structured Interviews

1. What is your overall impression of the system and its ability to support intelligence analysis and report writing?
2. If you were writing a real report for _____ (insert major here), would you want to use a system of this sort? Why or why not?
3. What were the most useful features of the system and why? (Specify for report writing)
4. What features were missing that would be helpful in intelligence analysis/ report writing?
5. Was there anything confusing, annoying, or difficult about using the system?
6. Do you have any other comments?

A.4 COVID-19 Consent Addendum

Additional COVID-19 Information Related to

Consent to Participate in In-Person Research at Virginia Tech

You have agreed to participate in a research study at Virginia Tech. The research study involves in-person contact or procedures. Here are some things you should know about in-person research while COVID-19 remains a risk:

Risks related to COVID-19:

If you choose to participate in this study, the risk for COVID-19 are about the same as those posed by similar activities while the virus is still spreading in your community. Similar activities could include grocery shopping, having your car repaired, or getting a haircut.

In addition, participation might increase risk to your family, the community, and the research team.

You should not participate if you have any conditions or risk factors that could make a COVID-19 infection more serious. Risk factors for severe illness include having other medical conditions such as asthma, diabetes, heart problems, or any other illness. Certain populations might also be at increased risk or unknown risk, including people aged 65 and older, people with disabilities, women who are pregnant or breastfeeding, people who are experiencing homelessness, and people who are part of racial and ethnic minority groups.

The information on people who need to take extra precautions is being updated regularly. We encourage you to check for the latest information before you decide whether to participate.

Please visit https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html for the most up to date information.

What we are doing to reduce risk to you:

Each lab or study has developed a process for conducting the research as safely as possible, given current knowledge about COVID-19. This process has been reviewed by the Human Research Protection Program at Virginia Tech. You will be given a sheet with information specific to your study. You should review this information and ask any questions before you agree to participate.

We will not conduct the study during times of increasing community spread or if we cannot obtain the necessary disinfecting supplies and equipment to reduce the risk of exposure.

Everyone working on the study has been instructed to stay home if they have any symptoms that could be related to COVID-19. If someone on the research team tests positive for COVID-19 and you have been exposed, someone will notify you. We will maintain a contact tracing log that is separate from your data and other details about your participation, and we will provide this log to the New River Health District (540 267-8240) who will conduct contact tracing in the case of a positive test. We will destroy this log 60 days after your last visit.

What you can do to reduce risk to us and to the community:

Do not participate if you have had any symptoms of COVID-19 in the past 14 days or have been in contact with someone who has symptoms. Symptoms include, but are not limited to, cough, shortness of breath or difficulty breathing, fever, chill, repeated shaking with chills, muscle pain, headache, sore throat, and new loss of taste or smell.

Do not participate if you have tested positive for COVID-19 in the past 21 days, even if you have not shown any symptoms.

Do not participate if you know you have been exposed to anyone who has tested positive for COVID-19 in the past 21 days.

Let us know if you test positive for COVID-19 within the next 14 days. We will provide your contact tracing log to university or public health authorities who will use the tracing log to contact others who may have been exposed during your visit.

Wash your hands frequently and observe current guidance on avoiding virus spread from the Centers for Disease Control.

Wear a mask or a cloth face covering over your nose and mouth. Depending on the study, another method may be used, such as physical distancing, a face shield, or Plexiglas barrier.

For the latest information on COVID-19 please visit: https://www.cdc.gov/coronavirus/2019-ncov/index.html

Signature Block

Your signature documents that you have read and understand the information outlined in this document and all of your questions have been answered. We will provide you with a signed copy of this consent addendum for your records.

Signature of Subject & Date

_____

Printed Name of Subject

_____

Signature of Person Obtaining Consent & Date

_____

Printed Name of Person Obtaining Consent

_____

A.5 IRB Approval Letter

VIRGINIA TECH

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE: October 16, 2020

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

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires October 29, 2024)

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

IRB NUMBER: **20-762**

Effective October 16, 2020, 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) 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:

***Please note: You must use COVID-19 Consent Addendum and provide it to participants ahead of time.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* 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

Exploring Professional Analyst Sensemaking Strategies in Immersive Space to Think

B.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 with Non-Novice Users (IRB #21-428)

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

Other Study Contact(s):

Kylie Davidson
Department of Computer Science, Virginia Tech
Email: kyliedavidson@vt.edu

Douglas A. Bowman
Department of Computer Science, Virginia Tech
Email: dbowman@vt.edu

Kirsten Whitley
Department of Defense

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 Professional Intelligence Analyst that is at least 18 years old who has normal or corrected (glasses or contacts) vision with strong English skills.

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 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 (90 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 (20 minutes, split into two separate sessions): During this phase we will perform a semi-structured interview on your experience in the daily study phase.

- Post-study Phase (20 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

Participant Rights: In addition to contacting the IRB, you may also contact Leslie Bell, Human Protections Director, at lebell@evoforge.org.

How many people will be studied?

We plan to include 40 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 Emerson 3 Building, and will consist of three 90-minute sessions for each participant split into 4 phases.

In the first "Pre-study" phase, you will be shown the experiment space, 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 an Intelligence Analysis training dataset 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 intelligence report data from a fictitious Intelligence Analysis training task/s. Using these sources, you will be asked to write a report that addresses what action or actions the terrorists are planning, where they will occur, and when they will occur and/or an outline of a paper. 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. After each session, we will conduct a semi-structured interview during the "Post-session" or "Post-study" phase. This interview will ask you about your experiences using the software. The "Post-study" 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 Risk)

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 many short documents from Intelligence Analysis tasks. We are preserving these documents in their original state, which may cause some emotional 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 insight into using VR/AR technologies for Intelligence Analysis.

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 partially sponsored by the National Science Foundation.

This research is partially 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.

Consent Script for Verbal Consent Agreement:

"Do you give the 3DI Interaction Group and InfoVis Lab at Virginia Tech permission to record my research participation through audio and screen video recordings?"

"Furthermore, do you consent to take part in this research?"

B.2 Questionnaires & Interviews

B.2.1 Pre-Study Questionnaire

1. Gender (Make only one)

o Male

o Female

o Other

o Prefer not to share

2. How are you feeling today?

_____

3. Do you wear glasses or contact lenses?

o Neither

o Glasses

o Contact Lenses

4. Handedness:

o Right-Handed

o Left-Handed

o Ambidextrous

5. Rate your fatigue level:

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

6. Rate your expertise with computers

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

7. How often do you use computers for work?

o Not at all

o Once a month

o Once a week

o Several Times per Week

o Daily

8. 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

9. How often do you play video games? (console or on the computer?)

o Not at all

o Once a month

o Once a week

o Several Times per Week

o Daily

10. How many times have you tried virtual or augmented reality?

o Never

o Once or Twice

o 3 - 10 Times

o More than 10 times

11. Comments:

_____

12. Rate your experience watching 3D stereo movies, in theaters and on TV:

1 2 3 4 5 6

Poor (Nausea, Headaches, etc.) o o o o o o Excellent

13. Comments:

_____

14. Cumulative years of intelligence analysis experience:

_____

15. During a sensemaking task, which of the following do you use? (An example of a sensemaking task is reading documents and writing a research paper)

o Highlights

o Annotations

o Note Taking

o Other _____

16. Please explain your traditional approach to writing a research paper that analyzes a set of text documents.

_____

_____

_____

_____

B.2.2 Condition 1 Interviews

Session 1

1. What is the overall strategy you used for analysis today?

2. Did your strategy for analysis change over the course of this session?

3. Describe to me the spatial layout you have formed thus far.

4. Did your spatial layout change over the course of this session?

Session 2

1. Did your strategy for analysis change over the course of this session?

2. Did the strategy you used today differ from the strategy you used during session one? If so, How?

3. Describe to me the spatial layout you have formed thus far.

4. Did the spatial layout you used today differ from the spatial layout you used during session one? If so, How?

5. Describe the transformation of your spatial layout to your report outline.

Session 3

1. Over the course of this analysis, how did you organize documents to extract meaningful information?

2. Does your spatial layout reflect the organization of your report? If so, How?

3. Did your spatial layout change over the course of multiple sessions, If so, How?

4. Do you see advantages to using a system like IST over traditional analysis approached for the sensemaking task?

B.2.3 Condition 2 Interviews

Session 1

1. What is the overall strategy you used for analysis today?

2. Did your strategy for analysis change over the course of this session?

3. Describe to me the spatial layout you have formed thus far.

4. How does the spatial structure you formed map to the outline you generated?

Session 2

1. Did your strategy for analysis change over the course of this session?

2. Did the strategy you used today differ from the strategy you used during session one? If so, How?

3. Describe to me the spatial layout you have formed thus far.

4. Did the spatial layout you used today differ from the spatial layout you used during session one? If so, How?

5. How does the spatial structure you formed map to the outline you generated?

Session 3

1. Did your strategy for analysis change over the course of this session?

2. Did the strategy you used today differ from the strategy you used during session one/session 2? If so, How?

3. Describe to me the spatial layout you have formed thus far.

4. Did the spatial layout you used today differ from the spatial layout you used during session one/two? If so, How?

5. How does the spatial structure you formed map to the outline you generated?

6. Do you see advantages to using a system like IST over traditional analysis approached for the sensemaking task?

B.3 Post-Study Semi-Structured Interviews

B.3.1 Condition 1

1. What is your overall impression of the system and its ability to support intelligence analysis and report writing.

2. If you were writing a real report, would you want to use a system of this sort? Why or why not?

3. If you were to use this system again, what would you do differently?

4. What were the most useful features of the system for analysis and why?

5. What were the most useful features of the system for report writing and why?

6. What features were missing that would be helpful in intelligence analysis/ report writing.

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

8. Do you have any other comments?

B.3.2 Condition 2

1. What is your overall impression of the system and its ability to support intelligence analysis and outline writing.

2. If you were writing a real report, would you want to use a system of this sort for your analysis? Why or why not?
3. What were the most useful features of the system for analysis and why?
4. Which strategy for analysis did you find to be the most effective?
5. Which spatial layout for analysis did you find to be the most effective?
6. If you were to use this system again, would you use a different strategy or spatial layout for your analysis? Why or why not?
7. Looking forward to adding automation, how could the system help you to develop an outline of your findings?
8. What features were missing that would be helpful in intelligence analysis?
9. Was there anything confusing, annoying, or difficult about using the system?
10. Do you have any other questions or comments?

B.4 IRB Approval Letter

VIRGINIA TECH

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE: June 16, 2021

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

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Immersive Space to Think with Non-Novice Users

IRB NUMBER: **21-428**

Effective June 16, 2021, 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) 4,6,7**
Protocol Approval Date:	**June 16, 2021**
Progress Review Date:	**June 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.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?
06/02/2021	PDCJRGZW	National Science Foundation (Title: IUCRC Pre-proposal Phase I Virginia Tech: Center for Space, High-performance, and Resilient Computing (SHREC))	Compared on 06/15/2021

* 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

Gathering Expert Feedback on Deeper Organizational Features

C.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 with Non-Novice Users 2023 (IRB #23-661)

Principle Investigator: Name : Doug Bowman
Department of Computer Science, Virginia Tech
Email: dbowman@vt.edu

Other Study Contact(s):

Kylie Davidson
Department of Computer Science, Virginia Tech
Email: kyliedavidson@vt.edu

Chris North
Department of Computer Science, Virginia Tech
Email: north@cs.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 Professional Intelligence Analyst that is at least 18 years old who has normal or corrected (glasses or contacts) vision

with strong English skills.

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 one 120-minute session in the following format:

- 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 (60 minutes): In this phase we will ask you to perform a task using the software shown in the Tutorial Phase.
- Post-study Phase (30): During this phase we will perform a semi-structured interview on your overall experience in the study phase with the goal of receiving feedback on system design and future directions for our prototype.

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 40 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 Emerson 3 Building, and will consist of a single 120-minute session consisting of 4 phases.

In the first "Pre-study" phase, you will be shown the experiment space, 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 an Intelligence Analysis training dataset 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 intelligence report data from a fictitious Intelligence Analysis training task/s. Using these sources, you will be asked to develop an understanding what action or actions the terrorists are planning, where they will occur, and when they will occur and/or an outline of a paper. 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. After the session, we will conduct a semi-structured interview during the "Post-study" phase.

This interview will ask you about your experiences using the software. The "Post-study" 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 Risk)

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 many short documents from Intelligence Analysis tasks. We are preserving these documents in their original state, which may cause some emotional 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 insight into using VR/AR technologies for Intelligence Analysis.

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 deidentified research records.

This research is partially sponsored by the National Science Foundation.

All subjects in the study do so of their own free will and a desire to further computing technologies.

Military and Civilian supervisors, officers, and others in the chain of command will not influence the decisions of their subordinated regarding participation in or consenting to this research.

Consent Script for Verbal Consent Agreement:

"Do you give the 3DI Interaction Group and InfoVis Lab at Virginia Tech permission to record my research participation through audio and screen video recordings?"

"Furthermore, do you consent to take part in this research?"

C.2 Questionnaires & Interviews

C.2.1 Pre-Study Questionnaire

1. Gender (Make only one)

o Male

o Female

o Other

o Prefer not to share

2. How are you feeling today?

---

3. Do you wear glasses or contact lenses?

o Neither

o Glasses

o Contact Lenses

4. Handedness:

o Right-Handed

o Left-Handed

o Ambidextrous

5. Rate your fatigue level:

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

6. Rate your expertise with computers

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

7. How often do you use computers for work?

o Not at all

o Once a month

o Once a week

o Several Times per Week

o Daily

8. 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

9. How often do you play video games? (console or on the computer?)

o Not at all

o Once a month

o Once a week

o Several Times per Week

o Daily

10. How many times have you tried virtual or augmented reality?

o Never

o Once or Twice

o 3 - 10 Times

o More than 10 times

11. Comments:

_____

12. Rate your experience watching 3D stereo movies, in theaters and on TV:

	1	2	3	4	5	6	
Poor (Nausea, Headaches, etc.)	o	o	o	o	o	o	Excellent

13. Comments:

_____

14. Cumulative years of intelligence analysis experience:

_____

15. During a sensemaking task, which of the following do you use? (An example of a sensemaking task is reading documents and writing a research paper)

o Highlights

o Annotations

o Note Taking

o Clustering

o Linking

o Other _____

16. Please explain your traditional approach to writing a research paper that analyzes a set of text documents.

_____

_____

_____

_____

C.3 Post-Study Semi-Structured Interviews

1. Can you describe the way in which you used the permanent (yellow) linking feature:
2. Can you describe to me the way in which you used the discovery (red) search linking feature:
3. How would you like to change the permanent linking feature to aid in your analysis?
4. How would you like to change the search linking feature (red) to aid in your analysis?
5. As you add more links into your analysis space, how, if at all, should we help you prevent clutter?
6. Do you see the linking feature interfering with your analysis process?
7. How do you see the linking feature benefiting your analysis process?
8. What do you wish that the linking features could do for you that it is not already able to do?
9. Can you describe to me the way in which you have used the clustering feature?
10. Can you describe to me the way you used the automated layout features?
11. How would you like to change the clustering feature to aid in your analysis?

12. As the number of clusters increases, how can we help to decrease clutter within your space?

13. Do you see a value in displaying a set of default clusters within the space?

14. Do you see the clustering interfering with your analysis process?

15. What do you wish the clustering feature could do for you that it is not already able to do?

C.4 IRB Approval Letter

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

VIRGINIA TECH

MEMORANDUM

DATE: July 25, 2023

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

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Immersive Space to Think With Non-Novice Users 2023

IRB NUMBER: **23-661**

Effective July 25, 2023, 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:	**June 28, 2023**
Progress Review Date:	**June 28, 2024**

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 July 25, 2023 updates consent form.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?
06/28/2023	PDCJRGZW	National Science Foundation (Title: IUCRC Pre-proposal Phase I Virginia Tech: Center for Space, High-performance, and Resilient Computing (SHREC))	Compared on 06/28/2023

* 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 D

Evaluating Organizational Feautres in Immersive Space to Think

D.1 Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
Consent to Take Part in A Research Study

Title of Research Study: : Evaluating Organizational Features in Immersive Space to Think (IRB #24-104)

Principle Investigator: Name : Doug Bowman
Department of Computer Science, Virginia Tech
Email: dbowman@vt.edu

Other Study Contact(s):

Kylie Davidson
Department of Computer Science, Virginia Tech
Email: kyliedavidson@vt.edu

Chris North
Department of Computer Science, Virginia Tech
Email: north@cs.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 affiliate that is at least 18 years old who has normal or corrected (glasses or contacts) vision with native

English-speaking skills.

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 various organizational features for immersive sensemaking with virtual reality (VR) and augmented reality (AR) systems. This research will help us to understand the affects of these organizational features on immersive sensemaking. 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 one 2.5 hour session split into the following phases:

- Pre-study Phase (20 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 (20 minutes): This phase will present the software we will be using as well as how the software will be used during the study. As completing an example task.
- Study Phase (1 hour 20 minutes): In this phase we will ask you to perform a series of four tasks using the software shown in the Tutorial Phase
- Post-study Phase (10 minutes): During this phase, we will perform a semi-structured interview on your 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 30-50 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 3DI lab (room 3215) of Gilbert Place and will consist of one 2.5 hour session for each participant split into 4 phases.

In the Pre-study phase, you will be shown the lab and given a brief overview of our research. You will then be asked to provide your informed consent to participate in the user study. Following the consent process, you will fill out a background questionnaire.

After that, we will go into the Tutorial phase. In this phase you will be shown the equipment we will be using for the experiment (Varjo XR-3 and the VR/AR controllers. Then we will introduce our software, called "The Immersive Space to Think (IST)." We will also give you a mock NYT Connections task 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 completing a series of 4 different NYT Connections games. Using the cards (documents) provided, you will be asked to sort the cards into categories of four cards. During each task, you will be provided with different organizational features to use to aid in your analysis. After finishing each task, you will submit your answer to the system and complete a series of questionnaires, and answer a brief semi-structured interview.

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.

Screen-video recordings and audio recordings will be taken during the study portion and the interview sessions of this 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 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 Risk)

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.

Will being in this study help me in any way?

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 NSF via the Center for Space High-Performance, and Resilient Computing (SHREC).

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.

All hardware used during this study (VR/AR headset, hand-held controllers, keyboard, etc.) will be wiped down with disinfecting wipes between each participant to help mitigate the spread of viruses.

Signature Block for Capable Adult

We are taking audio, screen, and video recordings of your participation in this research. Please initial the following statement to provide your permission for these recordings: I give the 3DI Interaction Group at Virginia Tech permission to record my research participation through audio, screen, and video recordings

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

Furthermore, your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

Signature of Subject: _____
Printed name of Subject: _____
Date: _____

Signature of person obtaining consent: _____
Printed name of person obtaining consent: _____
Date _____

D.2 Questionnaires & Interviews

D.2.1 Pre-Study Questionnaire

1. Gender (Make only one)

o Male

o Female

o Other

o Prefer not to share

2. Age

_____

3. Do you wear glasses or contact lenses?

o Neither

o Glasses

o Contact Lenses

4. Handedness:

o Right-Handed

o Left-Handed

o Ambidextrous

5. How many times have you tried virtual or augmented reality?

o Never

o Once or Twice

o 3 - 10 Times

o More than 10 times

6. Rate your quality of viewing experience while watching 3D stereo movies, in theaters and/or on TV:

1 2 3 4 5 6
Poor (Nausea, Headaches, etc.) o o o o o o Excellent

7. Have you ever played the NYT Connections Game?

o Yes

o No

8. If Yes, How frequently do you play the NYT Connections Game?

o Only once or Twice

o A Couple of times a month

o A couple of times a week

o Close to daily

o Daily

o I have never played the NYT Connections game

D.3 Post-Task Semi-Structured Interviews

1. How did you use the provided features in this condition to support your analysis?
2. Can you describe the strategy you used to solve the connections task in this condition?
3. Were the features in completing the task? And if so, why were they helpful?

D.4 Post-Study Semi-Structured Interviews

1. What are your thoughts on the _____ features and how they currently supports your sensemaking?
    - Freestyle
    - Clustering
    - Manual Linking
    - Discovery Linking
    - Clustering & Linking

2. What feature was most effective in assisting you in finding the connections?
3. Over the course of this study, what feature/s were the most effective in helping you stay organized during your analysis?
4. What condition best supported your strategy in solving the connections task?
5. If you were completing a sensemaking task for school (e.g., writing a research paper, related works, etc.) which features do you see yourself using and why?
6. Was there anything confusing, annoying, or difficult about using the system?

D.5 IRB Approval Letter

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

VIRGINIA TECH

MEMORANDUM

DATE: March 1, 2024

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

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Evaluating Organizational Features in Immersive Space to Think

IRB NUMBER: **24-104**

Effective February 29, 2024, 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:	**February 28, 2024**
Progress Review Date:	**February 28, 2025**

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:

A link to access the stamped documents can be found on the summary page for this protocol. Click on "View/Download Latest Stamped Forms" on the right side of the page or scroll to the bottom and click on "Click here to view/download latest stamped forms" to access the form.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* 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 E

Report Grading Rubrics

To develop the quality rubric we worked with a set of Professional Intelligence Analytics. Using an iterative development process we ended up with the following rubric with each question scored on a 6 point likert scale (Strongly Disagree: 0, Disagree: 1, Slightly Disagree: 2, Slightly Agree: 3, Agree: 4, and Strongly Agree: 5):

1. Concise:
   (a) The report was concise.
   (b) The report was written in a Bottom Line Up Front (BLUF) manner.
2. Compelling:
   (a) The report focuses on the facts.
   (b) The report effectively persuades the reader to take the threat/s seriously.
3. Clarity:
   (a) The report is clear.
   (b) The report uses plain English and avoids jargon.
   (c) The report's facts are expressed clearly.
   (d) The report is free of distracting punctuation and grammatical errors.
   (e) Abbreviations and capitalization's are correct and consistent.
4. Completeness:
   (a) The report is complete.
   (b) The report includes all the data facts and details needed to clearly follow the developments of the report's focus and support all conclusions.
5. Relevance:
   (a) The report contains only relevant information.
6. Bias:

   (a) The analyst remains objective in the presentation of the report
   (b) The argument being made is differentiated from the evidence being presented.
   (c) The analyst presents any uncertainty in events.

To develop the correctness rubric we use the ground truth document which is included with the training dataset. The rubric was broken down as follow:

1. Who:
   (a) Plotline 1 - 3 Actors (1 point each)
   (b) Plotline 2 - 3 Actors (1 point each)
   (c) Plotline 3 - 3 Actors (1 point each)
   (d) Coordinator Actor (3 points)
2. What:
   (a) Coordination Action (5 points)
3. When:
   (a) Date (3 points)
   (b) Time (2 points)
4. Where:
   (a) Plotline 1 Location (2 points)
   (b) Plotline 2 Location (2 points)
   (c) Plotline 3 location (2 points)
5. Other
   (a) Accurate Information and Claims (5 points)