Spaces to Think: A Comparison of Small, Large, and Immersive Displays for the Sensemaking Process

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

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

Leonardo Pavanatto‡
Center for Human-Computer Interaction
Department of Computer Science
Virginia Tech

Ibrahim A. Tahmid§
Center for Human-Computer Interaction
Department of Computer Science
Virginia Tech

Chris North¶
Sanghani Center
Department of Computer Science
Virginia Tech

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

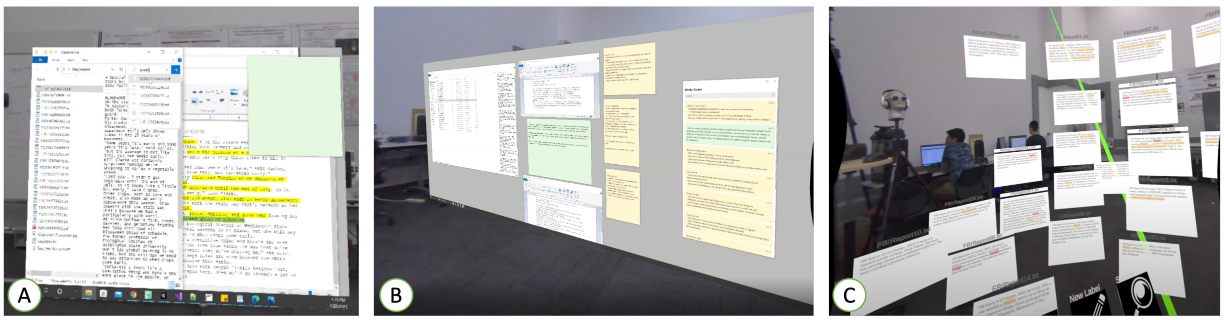

Figure 1: The conditions compared in this study. A: the small virtual monitor condition, which represented a 24" 4:3 monitor. B: the large virtual monitor condition, which represented a 4x2 array of 24" 4:3 monitors. C: Immersive Space to Think, where documents can be placed in 3D space surrounding the user.

Abstract

Analysts need to process large amounts of data in order to extract concepts, themes, and plans of action based upon their findings. Different display technologies offer varying levels of space and interaction methods that change the way users can process data using them. In a comparative study, we investigated how the use of single traditional monitor, a large, high-resolution two-dimensional monitor, and immersive three-dimensional space using the Immersive Space to Think approach impact the sensemaking process. We found that user satisfaction grows and frustration decreases as available space increases. We observed specific strategies users employ in the various conditions to assist with the processing of datasets. We also found an increased usage of spatial memory as space increased, which increases performance in artifact position recall tasks. In future systems supporting sensemaking, we recommend using display technologies that provide users with large amounts of space to organize information and analysis artifacts.

Index Terms: Human-centered computing—Visualization—Visualization techniques—; Human-centered computing—Visualization—Visualization design and evaluation methods Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

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

1 Introduction

Sensemaking tools have changed considerably over the past decade. We have traditional desktop scenarios involving features like windows, previews, and search bars, but these two dimensional (2D) representations are limited by screen space even when including "virtual" desktops that users can switch between [6]. The large, high-resolution displays of the Space to Think approach attempt to mitigate these spatial concerns, but it is still limited to flat 2D space that can only contain a fixed amount of information [2]. Newer display technology using immersive space through augmented reality (AR) or virtual reality (VR) has also been explored in a sensemaking context. Immersive analytics allows users to leverage expansive three-dimensional (3D) space to organize data and other analysis artifacts, potentially leading to better understanding through externalized memory and embodied interaction [7, 28].

One immersive approach, the Immersive Space to Think (IST), purports to assist users with understanding non-quantitative datasets through providing tools to organize, annotate, and synthesize findings [13]. The key way the approach differs from desktops or the Space to Think approach is that IST offers 3D immersive space as well as 3D interaction techniques, rather than 2D displays and 2D interaction techniques. Immersive space has been shown to provide benefits to both spatial memory [20, 37] and context switching [17]. Similarly, 3D interaction has been shown to provide benefits through embodied interaction [7, 11, 19]. However, it is not fully understood how well these benefits translate to the sensemaking process.

To better understand the tradeoffs and performance between these technologies, we conducted a within-subjects study to evaluate the benefits of immersive space and 3D user interaction techniques on the sensemaking process. Our conditions included a small monitor condition simulating the traditional desktop scenario, a large monitor condition simulating the Space to Think approach, and an immersive condition inspired by the IST approach.

We found that users preferred having more space while performing these tasks, as well as having greater engagement and interaction with documents as space grew. Participants leveraged 3D space by placing information in different depth layers to indicate importance. We also found evidence of users employing "working areas" in immersive space, which we define as physically separated areas that users devoted to a particular task or category, such as report writing or topic organization. In a similar vein, we found that increased amounts of space used in the IST condition had a weak correlation with performance, which should be explored in future studies. Lastly, we found evidence of increased spatial memory usage as space grew, with spatial memory being the most common strategy in a document search task in the IST condition.

These findings suggest that the IST approach is a promising way to understand large, complex, non-quantitative datasets that gives users ample amounts of space to assign meaning. It further provides benefits to users through spatial memory, allowing them to find documents or themes quickly through their organizational schema. Qualitative feedback from participants also suggests that future tools should move away from six-degree of freedom (6DOF) controllers for easier context switching and annotation. Overall, increased space provides increased user satisfaction and reduces frustration with the sensemaking task, and we recommend providing users with as much space as is feasible.

2 Related Work

2.1 Sensemaking & Immersive Analytics

While there have been many models created to describe the process of sensemaking, they all agree that it is a cognitively difficult process that is performed repetitively and iteratively where users "structure the unknown" [1, 21, 36, 39]. In particular, we focus on Pirolli and Card's model of sensemaking, which discusses how analysts perform tasks in two main loops: the foraging loop, and the sensemaking loop [36]. In the foraging loop, analysts find sources, skim them to identify topics, and lightly organize them by theme. In the sensemaking loop, analysts further extract data from the sources, create hypotheses and tell the story that lies beneath the sources. IST purports to assist analysts with the entire process by representing each source as a data artifact and allowing the analysts to organize these artifacts, annotate them, and synthesize their findings.

Visual analytics has helped analysts with the sensemaking process through providing tools to create interactive visualizations of abstract data to assist with understanding [2, 14, 16]. Several of these tools specifically support sensemaking with non-quantitative datasets, where analysts can leverage aspects of digital processing such as keyword search or semantic interaction to perform their tasks [4, 14, 29]. Immersive analytics is an extension of visual analytics that is combined with mixed reality to further visualize data in 3D immersive space [10, 31]. This field has been used to explore how immersive technologies can assist with decision making through the combination of human intelligence and machine intelligence and improving analytical capabilities [15, 41].

The IST approach in particular has been explored as a way to better understand large, complex non-quantitative datasets in an immersive environment [25]. Lisle et al. observed users creating 2.5-dimensional structures and that document display order affected how those documents were processed [26]. Davidson et. al looked at how users processed data over multiple sessions. They found that users refine their existing structures over time, as well as evidence of a linear transformation path between artifact organization and report synthesis [13]. Tahmid et al. explored tools to assist with IST organization by creating automated and semi-automated clustering interaction techniques and found that participants preferred and performed better with the semi-automated technique [42].

Other studies on non-quantitative immersive sensemaking looked at how people interact with data artifacts and the environment around them. Luo et al. found that single users and groups of users would leverage furniture in AR settings to scaffold their arguments and create meaning [28]. Zhang et al. proposed leveraging a users' interaction history in order to support their recall, annotation, and sharing of identified concepts and themes [49]. They further designed prototype interaction methods using their proposed criteria to support sensemaking in immersive analytics. These tools and findings can be used to design better immersive analytics approaches and applications. However, the critical question of whether immersive analytics provides measurable benefits over non-immersive approaches remains unanswered.

2.2 Comparison of Displays

There have been several studies that have looked at how different display technologies, such as AR, VR, or large high-resolution 2D displays, impact how humans process information [3, 8, 22, 32, 35]. Furthermore, while we focus on immersive analytics with AR/VR HWDs, large, high-resolution displays offer many benefits as a competing option. For example, wall displays are semi-immersive in that it takes effort to look away from the screen [8, 24]. The Space to Think concept, as another example, affords the ability to sort large amounts of documents in 2D space [2]. Andrews et al. found that users could leverage spatial memory to recall documents more easily while using these displays [3]. IST furthers these advantages, by providing a both surrounding display and 3D depth that provides more space for users.

Often, display technologies are compared against each other to ascertain relative strengths and weaknesses. Czerwinski et al. looked at user satisfaction and performance between small and large 2D displays, finding users performed tasks faster, had higher recall, and preferred the larger display [12]. Cetin et al. continued these comparisons, looking at visual analytics task performance with small and large displays and also varying display resolution [9]. They found that different analysis techniques and performance were found between the small and large conditions, but resolution did not impact task performance.

Other comparative studies more directly evaluate the effect of display style on performance metrics. For example, Andrews et al. looked at the effect spatial memory had on Space to Think and its large, high-resolution displays, and found that smaller displays required more effort to recall where certain virtual artifacts were kept [3]. Ping et al. designed a depth perception task for both VR and AR and compared users' performance in the two modalities, finding AR afforded better depth estimations [35]. In another comparison, Park & Kim looked at how motivation changed usage patterns in AR and VR for users who were shopping for clothes, where different motivating factors impacted purchase patterns [33]. Voit et al. performed an empirical evaluation study of smart artifacts using five different methods including in-situ, lab studies, AR simulation, virtual reality simulation, and online surveys [45]. These studies show that the differences between display technology can change how people perform tasks.

2.3 Spatial Memory & Organization

Robertson et al. define spatial memory as "the ability to remember where you put something," and this can be supported digitally in many ways [38]. Andrews and North explored spatial memory and their Space to Think concept, stating that this effect occurs through embodied cognition and that natural human spatial abilities afford this form of distributed cognition [5]. They discuss that the effect wasn't limited to their large 2D displays, but likely also applied to other forms of immersive display and environments like the ones we use for IST.

Landmarks are a key assistive tool in spatial memory that can be created easily in immersive environments. Wickens and Hollands argue that designers should synthetically create landmarks to support navigation and spatial memory tasks [46]. Through the usage of these landmarks, immersive environments can assist users in finding and revisiting digital objects in terms of speed and accuracy [30,38]. Furthermore, in a study by Liu et al., when documents were evenly spaced in cylindrical layouts, participants performed worse at spatial memory tasks than when in flat or semicircular layouts [27]. Their cylindrical layouts lacked any landmark cues for the participants to latch onto and assist their spatial memory, which may have caused some amount of the performance disparities. IST using an AR display naturally supports landmarks by allowing the user to leverage real-world objects as their anchoring features.

Egocentric body movements, as well as interaction techniques, have also been shown to assist with spatial memory [37,43]. Rädle et al. performed two studies evaluating a pan and zoom interface on a large, high-resolution display using egocentric body movements as compared to traditional multi-touch panning and zooming [37]. They found that their new technique was faster and more efficient than the traditional method. Furthermore, participants who used their new technique performed significantly better on a long-term spatial memory task. Similarly, Tan et al. found egocentric interfaces that leveraged the awareness of the body's position relative to its environment assisted with spatial memory tasks [43]. As our design leverages several different egocentric body movements to interact with data artifact, we should see a benefit for spatial memory.

3 Immersive Space to Think Prototype

The IST approach purports to assist users in understanding large, complex, non-quantitative datasets through the ability to organize, annotate, and present data artifacts in 3D immersive space [13,25,26]. IST makes use of a tracked VR/AR head-worn display (HWD), tracked handheld controllers, and a tracked standard keyboard for entering text. IST users can view documents, images, and other data artifacts as virtual objects in an immersive 3D space, can place those artifacts anywhere in the space, and can create annotations such as highlights, notes, and labels. Examples can be seen in Figure 8.

Since we wanted to directly compare IST to traditional 2D display technologies, we needed to control confounding variables such as interface features and ergonomics as much as possible. Therefore, while we kept most of the features seen in previous studies using the IST approach, we have adapted it in several ways. More on how we controlled confounding variables can be seen in Section 4.3. Features unique to our IST prototype are listed below.

Seated Position - Traditional 2D displays are typically used while seated. Therefore, we adapted IST to use a rotating chair with a desk attachment that can rotate around the user. This allows the user to face any direction while organizing their thoughts in 3D immersive space and retain the ability to annotate documents using a keyboard.

Automatic Rotation - Previous work using IST found participants had difficulty rotating documents [26]. Since we were using a seated position, we decided to automatically rotate all documents to face the user, which improves readability and reduces the required effort during document placement.

File Browser - In our 2D conditions, users would be accessing documents through the built-in Windows file browser. We integrated this concept into IST through our own file browser that uses different icons to represent text and images, and allows users to preview the contents of a file. This can be seen in in Figure 8A.

Report Generation - As part of our experimental design included later stages of the sensemaking loop where users synthesize their understanding [36], we included a special note object in which users could write their "report." This was identical to other notes, except it appeared as a gray document to make it visually distinct.

4 Experimental Design

4.1 Research Questions

The goal of our study was to understand what benefits the IST approach provides to analysts compared to traditional 2D display technologies and how larger amounts of space aid sensemaking. We designed the study to address four research questions.

RQ1: *How does an increase in available space affect how users process large datasets?* As the available space increases, users can leverage that space to avoid the additional cognitive overhead required by the management of overlapping windows. We hypothesize that as the space increases, users will require less spatial management and can allocate those resources to solving sensemaking tasks.

RQ2: *As available space increases, how do users utilize the additional space?* Large datasets can fill a workspace quickly. But how do users manage those spaces when they have different amounts of space? How users manage to get around the disadvantages of the smaller spaces as well as how they utilize the larger spaces can affect how they perform sensemaking tasks. We hypothesize that we will see different strategies that users employ to manage the space in each condition. Furthermore, we hypothesize that users will leverage spatial memory more as space increases, as they can assign meaning to different locations.

RQ3: *How do 3D immersive space and 3D interaction methods impact user sensemaking strategies as compared to traditional 2D displays?* Two of the key ways IST is a novel approach for sensemaking tasks are that it leverages 3D immersive space and 3D interaction methods. How does the ability to use 3D depth and 3D interaction affect the user experience? We hypothesize that users will make use of depth to organize documents in ways that assist their spatial memory and writing process, but also that 2D interaction will be preferred due to familiarity and simplicity.

RQ4: *What, if any, is the comparative difference in performance between the conditions, with their differing amounts of space?* Naturally, the best indication of a tool's worth is how well it performs the task it was designed for. We also recognize that there are many factors that go into the process of sensemaking, so it may be difficult to measure significant differences in performance, as in many studies about the similar process of learning [40,48]. However, we hypothesize that we will see performance increase as available space increases due to reduced cognitive overhead in managing the space.

4.2 Conditions

To address our research questions, we designed three conditions varying the amount and type of space provided to the participants (see Figure 1). To remove confounding variables such as HWD weight or clarity of images, all conditions were designed to use the same HWD hardware. This meant that for the 2D conditions, we used virtual displays, which have been explored as a replacement to physical displays [34]. The first condition, *small*, simulated a 2D 24-inch display with a 4:3 aspect ratio and resolution of 1600x1200 pixels, and used 2D interaction via a mouse and keyboard. The second 2D condition, *large*, simulated a 4x2 array of the small displays (with no boundaries or bezels between the monitors) with a total size of 6400x2400 pixels, similar to the Space to Think concept [2]; it used the same mouse and keyboard as *small*. Both *small* and *large* used the Windows 10 operating system and its Windows Explorer (for file browsing), WordPad (for text document display and markup), and Sticky Notes (for note taking and labels) applications. The last condition, *IST*, used 3D immersive space with 3D interaction methods based on a handheld tracked controller. The *IST* condition ran custom software developed in the Unity game engine, with the features described in Section 3.

4.3 Apparatus & Experimental Setting

All conditions used the Varjo XR-3 HWD, which has a resolution of 1920x1920 pixels per eye and a field of view of 115 degrees [44]. We used Valve's SteamVR 2.0 tracking system for HWD and device tracking. In the *small* and *large* conditions, users viewed virtual monitors through the HWD. All conditions used the XR-3 as a video see-through AR display; users viewed the virtual displays and documents overlaid on a 12-megapixel color video view of the real world. The XR-3 ran on a desktop PC with an Intel Core i9-9800 processor and an NVIDIA GTX 2080 graphics card. Participants sat in a rotating chair with a desk attachment such that they could easily access the keyboard. This arrangement also helped with the *large* condition, as the virtual display was larger than the XR-3's field of view. The rotating desk afforded the participant the ability to face any part of that display and still have easy access to typing. Furthermore, for the *small* and *large* conditions the participants were given a Logitech K380 TKL Wireless Scissor Keyboard and a Logitech Performance MX Wireless Mouse to use with the Windows 10 operating system. In the *IST* condition, participants used a Valve Index wireless controller, a VIVE Tracker 2.0 to track the desk position, and a Logitech K780 full size Wireless Scissor Keyboard (we used the number pad enter key–which was not available on the other keyboard–as a way to confirm text entry in the IST condition).

4.4 Experimental Tasks

One aim of this study was to understand how users utilize the space given to them as the available space increases. Therefore, our datasets needed to be large enough to take up all the space available in the large 2D condition, ensuring that they had to make some tradeoff decisions in the two 2D conditions. To this end, we used three datasets that were created as training sets for intelligence analysts. These were "The Sign of the Crescent" (crescent), "The Case of Wigmore Vs. Al-Qaeda" (manpad), and "Stegosaurus (Excellent Apples)" (stegosaurus). Crescent and manpad were developed by the intelligence community to assist with intelligence analyst training, and stegosaurus is the 2006 VAST challenge dataset. These datasets have been used in prior studies on sensemaking tools [13, 50], and Wu et al. compared crescent and manpad to other VAST challenge datasets [47]. We used 24 documents from each dataset (while ensuring that all relevant documents for completing the task were included), such that the small condition could not display all the documents at once, the large condition would have difficulty displaying all documents, and IST could easily fit all the documents. This also ensured that each task was of similar size to each other. In addition, we kept the dataset order constant while using a latin-square rotation for the condition, thereby decoupling condition and dataset.

These datasets were intended to be challenging and require close reading of the documents in order to connect people, places, and themes to solve the intelligence analysis task. This allowed us to simulate a real-world sensemaking task that required effort from our participants. Participants were asked to spend at least 30 minutes with the dataset before submitting their report to ensure they were using a sufficient amount of effort on the task. We asked them to report on four main questions: *who* are the actors in the plot; *what* are they planning on doing; *when* are they planning on executing the plot; and *where* is the target(s) or location(s)?

4.5 Procedure

This study had four phases: a pre-study phase where participants were provided an introduction to the study and filled out a background questionnaire, three main session phases (performed at separate times, with a minimum time between sessions of one hour), and a post-experiment phase. The main sessions were further divided into a training phase, a study phase, and a post-study phase. These are described in detail below.

During the main session phases, participants were given a tutorial for the interface and condition they were using that session. The order of conditions was counterbalanced, while the dataset order remained constant. Each participant was given crescent first, then manpad, then stegosaurus to solve. The tutorial for the small and large conditions instructed participants how to use the Windows operating system, Microsoft WordPad, and Microsoft Sticky Notes. With each app, they were shown how to create, move, and resize windows, select text, and edit text where applicable. In addition, they were instructed on how to use the preview section of the file browser to see file contents as well as how to search the document set for keywords using the file explorer. For IST, participants were instructed how to open the file browser and open the documents. They were further shown how to move documents, scroll documents, select text, highlight text, and copy text. They were then shown how to create notes, labels, and their final report document and how to edit each of them. Lastly, the participants were introduced to the search feature and how to find documents (including open and unopened documents) that contained the search string. Participants were given as much time to explore the features as needed. Each tutorial used an example dataset of articles taken from the CNN.com website. For the main session study phase participants were given an introduction to the dataset they were going to analyze. As soon as they indicated they were ready, they were given the dataset and the experimenter started the recordings. This subphase took a variable amount of time as people perform the sensemaking task at different speeds and participants self-reported when they were finished. Then, the post-session subphase was split into two parts: a semi-structured interview performed while still in AR, and the UEQ and NASA-TLX questionnaires. The semi-structured interview aimed to gauge how the condition assisted their performance. The questions asked were: *Please walk me through how you analyzed the dataset in this session. How did the available space impact your analysis? I'm going to ask you to find a particular document in the dataset. You may use any features available in the main phase to find it. Once found, please read the title of the document. Can you please find [description of document in the session's dataset]? Could you please describe how you located that document? Could you please walk me through how you reported your findings?* Each main phase took 60-120 minutes.

After the completion of all three main sessions, participants were given another semi-structured interview to describe their experiences with each condition in the post-experiment phase. We asked the following questions: *Could you please compare and contrast the three conditions you experienced over the course of this study? Which condition, in your estimation, best supported the task and why? Please rank the other two as well. As the amount of available space changed, how did you organize documents? How, in particular, did 3D space in the IST condition impact your process? Was there anything confusing, annoying, or difficult about the interface in each condition*? This interview took 10-15 minutes. Participants were compensated with $70 for an expected 5-6 hours of work.

4.6 Data Collection & Measures

We collected data for this mixed methods study in a number of ways. We administered questionnaires using Google Forms, including the User Experience Questionnaire (UEQ) [23], and the NASA Task Load Index (NASA-TLX) [18]. Log files were generated from each condition logging interactions with windows (such as grabbing or maximizing) in the small and large conditions, and IST interaction methods in the IST condition. Furthermore, camera position and orientation was recorded up to ten times a second in all conditions. In the IST condition, the controller position and orientation was also recorded. The Varjo Base software made video recordings of all conditions, and the interviews were recorded using Apple's iPhone Voice Memos app. Lastly, all final layouts were recorded using the Sticky Notes app's database for the small and large conditions, and a save file for the IST condition. All data was stored on Google Drive.

4.7 Participants

Our participant pool was restricted to university students who had strong English skills and were eighteen years old or older. Eighteen people were recruited through a human-computer interaction email list. However, two participants did not return for later sessions. Four were removed for not putting in the requisite effort of 30 minutes

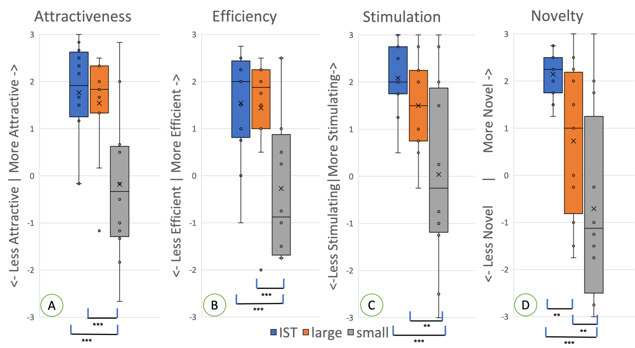

Figure 2: UEQ results for the aggregate categories of Attractiveness (A), Efficiency (B), Stimulation (C), and Novelty (D).

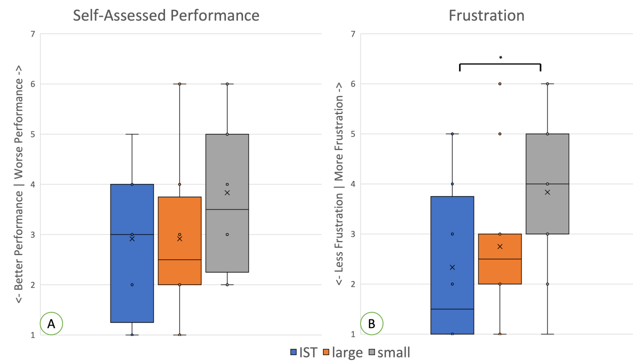

Figure 3: NASA-TLX results for self-assessed performance (A) and frustration (B).

and/or writing a minimum two-paragraph report. The remaining twelve participants (five female) had a mean age of 23.7 and a standard deviation of 3.33. Two wore glasses, and two wore contact lenses. All participants had prior experience with AR or VR. The study was approved by the institution's instructional review board.

5 Results & Discussion

To analyze the effects of our independent variables on our dependent measures, we performed Analysis of Variance (ANOVA) tests, with Tukey's post-hoc tests with Bonferroni corrections to determine pairwise significance. We used Kruskal-Wallis for categorical-continuous correlation comparisons, and Pearson for continuous-continuous comparisons. In all plots, significance is represented by asterisks: a single asterisk represents $p < 0.05$, double asterisks represent $p < 0.01$, and triple asterisks represent $p < 0.001$.

5.1 User Experience & Workload

Participant perceptions of user experience assist with our understanding of the differences between the three conditions to help answer RQ1. We performed ANOVAs to see the effect of condition on the aggregate category ratings in the UEQ and found significant differences in attractiveness ($F(2,22) = 12.16, p < 0.001$), efficiency ($F(2,22) = 10.79, p < 0.001$), stimulation ($F(2,22) = 11.64, p < 0.001$), and novelty ($F(2,22) = 23.55, p < 0.001$). Boxplots of these measures can be seen in Figure 2. For attractiveness, post-hoc analysis revealed significant differences between the IST and small conditions ($p < 0.001$), and the large and small conditions ($p < 0.001$). We found the same significant differences for efficiency ($p < 0.001$ for both pairs), and stimulation ($p < 0.001$ for IST v. small and $p = 0.00246$ for large v. small). These results demonstrate the small condition provides an undesirable user experience, while both the large and IST conditions, with ample space to view and organize documents, provide a high-quality experience.

The novelty subscale of the UEQ showed a slightly different pattern of significant differences, with post-hoc analysis showing that IST was considered more novel than either the large ($p = 0.00197$) or small conditions ($p < 0.001$), while large was more novel than the small condition ($p = 0.00164$). These results indicate that the small condition feels ordinary, while large has a higher degree of novelty (but note the range of opinions on this), and IST is considered highly novel by almost all users. AR and immersive environments are still an uncommon experience, which can contribute to a satisfying user experience. However, the difference between the large and small conditions was surprising. It indicates that this much seamless space without bezels or multiple monitors is itself novel; today's users are used to having smaller and more constrained display spaces.

In the NASA-TLX results, we found trends or significance in both self-assessed performance and frustration, and boxplots of these metrics can be seen in Figure 3. Self-assessed performance was trending towards significant ($F(2,22) = 2.584, p = 0.0981$). Post-hoc analysis revealed weak trends suggesting better perceived performance with IST than with small ($p = 0.147$) as well as with large as compared to small ($p = 0.147$). Furthermore, we found significance in the frustration metric ($F(2,22) = 4.363, p = 0.0254$). Post-hoc analysis found significantly more frustration with small than with IST ($p = 0.0127$) and a weak trend between small and large ($p = 0.116$). These results suggest that participants found the lack of space in the small condition to affect their frustration and ability with solving the task, while large and IST provided enough space for sensemaking with the datasets.

5.2 Qualitative Feedback

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

In particular, we asked participants to rank the conditions based on how well each of them supported the experimental task of sensemaking for intelligence analysis. Seven participants (58%) ranked IST first, while five (42%) ranked large first, and no participants preferred small. Participants who ranked large higher than IST still had positive things to say about IST: *[IST] had enough space to cluster different files, which was not the case for [large],* said P1, and P3 stated *[IST] was just more fun.* P6 went further, saying that organization was better in IST: *I loved using all the space in IST, and you could put things in their own section, and categorize them and you can move the panels around as you see fit. With the monitors it was more limited, you didn't have the comfort room.*

Participants who preferred IST over large praised IST for assisting with their organization. P17 stated *I could organize my notes a lot better... but I was able to swivel away from them when I didn't want to look at them ... that was better than minimizing the notes.* P15 stated that IST assisted with their spatial memory: *In IST I had so much space ... I had the space to visualize what I was working with and place the documents the way I wanted to and knew where I could find them compared to the other screens where you couldn't put 15-20 documents on one screen with your notes.* Participants felt that the amount of space and ability to freely organize documents was a particular strength of IST.

Some participants discussed issues they had with IST. Three participants raised the issue of device switching from interacting with documents to typing notes. For example, P13 said that the issues with the controller hampered their ability to offload cognition, stating *in [IST] I felt I took notes the worst because I kept having to pick up and put down the controller in order to write down things.* Participants who ranked large higher than IST often did so because of such minor UX issues, which is to be expected with a research

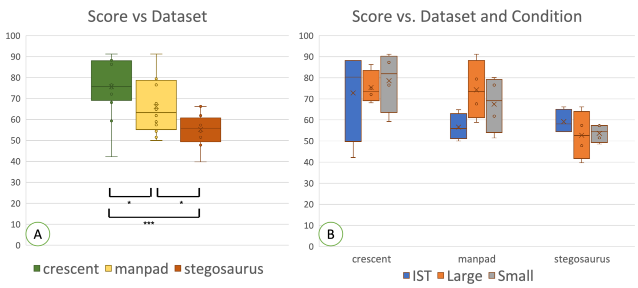

Figure 4: A: report score by dataset. B: report score by dataset and condition.

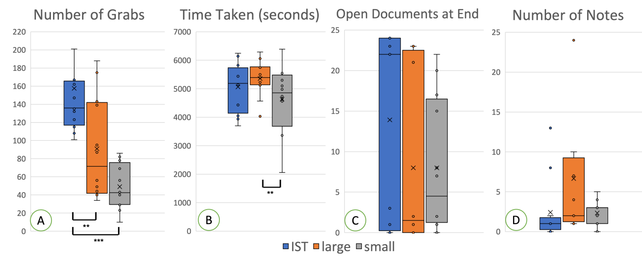

Figure 5: Effects of condition on user strategies. A: number of grab interactions in each condition B: time taken in each condition. C: number of documents that remained open at the end of the session for each condition. D: number of notes created in each condition.

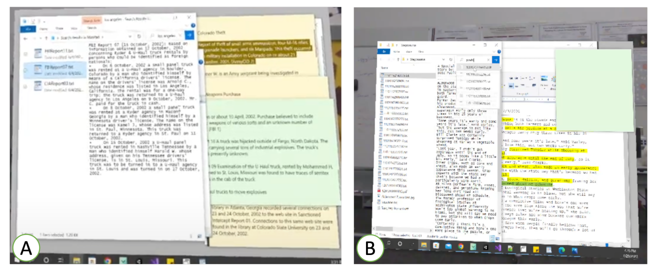

Figure 6: Examples of strategies seen in the small condition. A: P9 utilized the document preview on the file browser instead of opening files. B: P2 had to stack windows on top of each other during the task.

prototype. P6 stated that *I think IST would be above large with just a few minor improvements,* and P10 said *the gap between large and IST is very small.*

Participants discussed how the use of 3D depth in IST narrowed their focus to particular documents. P9 felt depth could signify the importance of documents: *for IST you could have things closer and farther, instead of just 2D spreading things out ... Things that weren't as important could be really small, having it further away made it obvious that it was not significant at the moment.* Some users also felt that it gave them more space for organization, and P17 even stated that IST changed how they worked: *If I was trying to do the same thing on my own computer I would have done something similar to what I did with the large monitor, but with IST I completely changed how I operate and I could have a bunch of things open and it not be cluttered. I don't like having too many things open because I don't like clutter, but I was able to contain documents to their own space.*

Participants also discussed how IST afforded the use of spatial memory during sensemaking. P4 used their spatial memory as a key aspect of their organization *The IST was amazing because I had the 360 degree space to organize windows. I actually made a good arrangement of my windows and I knew where they were.* Participants could leverage the environment itself as a way of understanding the document set as a whole.

5.3 Task Performance

The first author evaluated each participant's report for accuracy in answering the four questions outlined in the instructions: who, what, where, and when. Equal weight was given to each question, and when there were multiple answers to a question, equal weight was assigned to each answer. We performed an ANOVA to test whether condition affected report score, and found no significant effects ($F(2,28) = 0.375, p = 0.691$). However, a separate ANOVA with dataset as an independent variable found that dataset significantly affected score ($F(2,22) = 14.88, p < 0.001$). Post-hoc analysis further found significance between crescent and manpad ($p = 0.0343$), crescent and stegosaurus ($p < 0.001$), and manpad and stegosaurus ($p = 0.0104$). We ran a two-factor ANOVA looking at the effects of dataset and condition on score and found no significant main effects and no significant interaction between the two factors. Boxplots showing these results can be seen in Figure 4. These findings suggest that condition was not the only factor determining sensemaking performance (which is typical in studies of complex tasks). We do note that on the dataset that was apparently the most difficult (stegosaurus), IST had the highest mean and median scores.

5.4 User Strategies

Figure 5 shows comparisons of user interactions across conditions. We looked into how each condition affected the number of times participants grabbed documents in order to move them in the available space and found significance ($F(2,22) = 14.05, p < 0.001$). Post-hoc analysis revealed that IST had significantly more grabs than large ($p = 0.00373$) and small ($p < 0.001$). We posit this is due not only to the additional space affording more organization, but also to participants using spatial organization of documents and notes as a primary strategy for sensemaking in IST (see Figure 8), as suggested by participant interviews in Section 5.2.

We investigated how much time participants spent performing the tasks in each condition. As our tasks were difficult, participants spent a mean of 5018 seconds (standard deviation of 903 seconds), which was 83.6 minutes. We compared how long each participant took in each condition, and found a significant effect ($F(2,22) = 4.366, p = 0.0253$). Post-hoc analysis found a significant difference between small and large ($p = 0.00991$), with participants taking significantly less time in the small condition. Combining this with the increased frustration found for the small condition implies that participants wanted to be done with the small condition as soon as they could be. We further ran a correlation test between time taken and score, but found no significance between the two measures.

One aspect of strategy that differed among various sessions was whether participants decided to keep documents open to the end of their session or to close documents as soon as they read them. While there was only a trend for an effect of condition on the number of documents open at the end of a session ($F(2,22) = 2.677, p = 0.0911$), post-hoc analysis revealed weak effects between IST and large as well as IST and small ($p = 0.135$ for both), with participants keeping more documents open in IST than either small or large (Figure 5). While further investigation is required, this implies that participants felt more comfortable leaving documents open and organized into layouts in the IST condition, which may be due to the increased space that afforded simultaneous viewing.

We also looked at the number of notes generated in each condition and found weak significance ($F(2,22) = 3.416, p = 0.0511$). Post-hoc analysis revealed trends for significant differences between the large and IST conditions ($p = 0.075$) as well as the large and small conditions ($p = 0.0669$), where they made more notes in large than

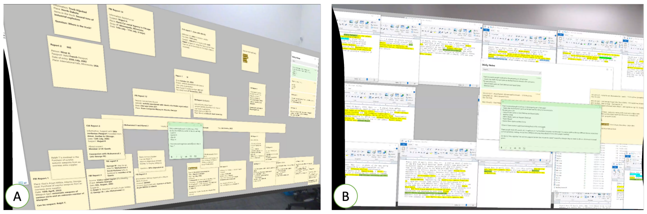

Figure 7: Examples of strategies seen in the large condition. A: P4 created a note for each document in their dataset and organized them into clusters. B: P2 kept all documents open and arranged around the screen so they could refer to them while writing the report.

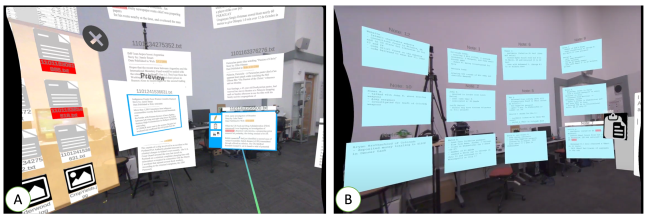

Figure 8: Examples of strategies employed in the IST condition. A: P13 used timelines, where they highlighted the dates in each document and organized them according to the date. B: P17 created a wall of notes that they referred to (instead of the document set) during the report writing synthesis.

either other condition (Figure 5). An example can be seen in Figure 7A. We believe this could be caused by a combination of factors. First, the small condition simply didn't have enough space to support many notes, with P7 stating *As I had more space, I had more sticky notes. In [small] I had 1 sticky note with all my information on it.* Furthermore, the large condition couldn't support both viewing all documents simultaneously and having multiple notes. Since sticky notes use less space, some participants preferred them over open documents, as P17 notes: *[In the large condition] I could open multiple documents and see them all at once, but I couldn't do that and also have my sticky notes open and arranged.* Finally, as noted above, there were minor issues with switching between document interaction and typing that prevented more notes being written in IST. Additionally, we ran a correlation test between number of notes and score and found no significance.

Participants had to create ways of dealing with the lack of space in the small condition. Five of the twelve participants did not open documents in WordPad. Instead, they elected to read the documents in the preview section of the file browser window which reduced occlusion but only allowed viewing of one document at a time. An example can be seen in Figure 6A.

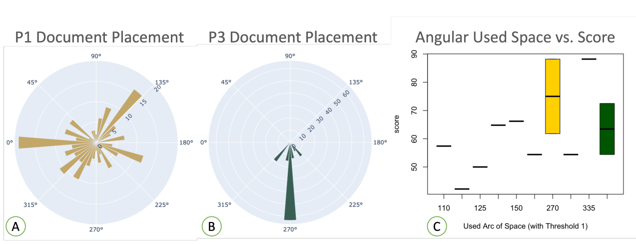

Figure 9: Participants varied how much space they utilized in the IST condition. A: P1 used a large arc of space for placing documents. B: P3 used a small arc of space for organization. C: comparison of the amount of angular used space (with 5-degree bins) and report score.

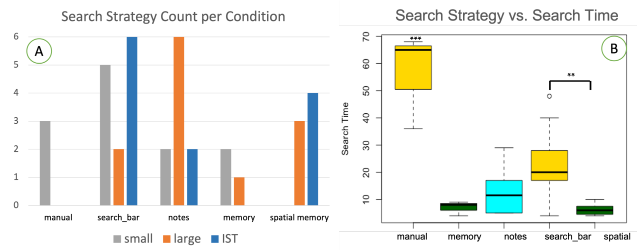

Figure 10: Participants used several different strategies in order to re-find documents after the sensemaking task. A: number of times each strategy was used in each condition. B: how long it took users to find documents with each strategy.

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

We investigated the use of space in IST further, by looking at locations where data artifacts were placed relative to the center-point of their work area (where their head was located in space). We sorted these into five-degree bins relative to their head's yaw and considered a bin to be "used space" if at least two artifacts were placed there at any time during the session. We found wide variation in the amount of angular used space. Two examples can be seen in Figure 9. A correlation test found a weak correlation between the amount of angular used space and score ($r(10) = 1.74, p = .112$). Working areas and using large amounts of angular space are potentially useful strategies in IST that should be explored in future studies.

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

5.5 Re-finding Strategies

One aspect of 3D immersive space is that it affords the ability to rely on spatial memory [6, 8]. We wanted to see whether this would improve participants' ability to find documents after the sensemaking session. In our post-session interviews, we asked participants to find a particular document by giving a brief description. We were able to categorize five different types of re-finding strategies participants employed: manual, search bar, notes, memory, and spatial memory. We defined *manual search* as when the participant would use the preview feature of Windows or IST to go through files one by one to find the document. *Search bar*, in contrast, was when the participant used the search feature in any condition. The *notes* category was defined as when participants referred to their notes to find the document. The *memory* strategy was when the participants simply recalled the document without referring to the interface at all. Lastly, *spatial memory* was defined as when the participant remembered where they had organized the document in space and referenced it at that location. A bar chart of counts for each re-finding strategy in each condition can be seen in Figure 10. We can observe that the spatial memory strategy could only be employed in the IST and large conditions, and that users in the small condition were primarily limited to using the search bar and manual strategies.

We measured the time to re-find the document. The timing results for each category of search can be seen in Figure 10. An ANOVA revealed a weak effect of condition on re-finding time ($F(2,22) = 3.292, P = 0.0561$). Post-hoc analysis found significant differences between small and large ($p = 0.0378$). We believe this is mostly due to the manual search strategy only appearing in the small condition. The lack of space seemingly affected the organization of participants' sensemaking process that they resorted to more inefficient search strategies. We ran an ANOVA to look for effects of re-finding strategy on the time it took to find the document and found significance ($F(4,7) = 558, p = 0.0242$). Post-hoc analysis revealed significant differences between manual and all other strategies ($p < 0.001$ in each case), and between search bar and spatial memory ($p = 0.00805$). This supports our supposition that small performed worse than IST or large in this task due to the manual search strategy. Furthermore, spatial memory was one of the best strategies (similar in efficiency to the memory strategy), and is only possible with larger amounts of space, which led to 33% of IST participants using this strategy without being prompted.

5.6 Key Takeaways

In RQ1, we asked how amount of space affects users' sensemaking. We found that users can become frustrated with inadequate space for sensemaking, as evidenced by the UEQ and NASA-TLX feedback. Participants found the small condition significantly less attractive, efficient, and stimulating than either other condition. Participants spent significantly less time in the *small*, which we believe was due to their frustration with the lack of space. We further found that as space increases, so too does the amount of interaction users have with documents, seen particularly with how often participants moved documents around the space. This was somewhat at odds with our hypothesis that larger spaces would require less spatial window management; it seems that when space became too small, users simply avoided using space to organize information.

For RQ2, we found that strategies varied in each condition as participants had more space to work with. In particular, we saw participants in *small* develop the strategy of using the preview feature in Windows to conserve space. This meant that they were willingly sacrificing the ability to use other features, such as highlighting or organization techniques. In contrast, we found participants making more notes in *large* than the other conditions. We believe this is an artifact of them having more space, but not quite enough to keep all the larger document windows open. Furthermore, we saw IST provide users with the ability to keep all the documents open at the same time. In addition, we found several participants assigning meaning to spaces while using IST. Similarly, we found that participants relied on their spatial memory more when given more space, with more instances of leveraging spatial memory seen in IST than *large*, and none at all in *small*.

For RQ3, we observed that 3D interaction affected where participants placed documents. For example, participants liked bringing documents closer to themselves when interacting with them directly, which allowed them to focus on particular documents. Furthermore, participants discussed how 3D interactions assisted with their spatial memory and their recall of concepts or themes.

Finally RQ4 asked how performance varies with amount of space. There was no significant difference in performance between the three conditions. However, within IST in particular, we found a weak trend correlating performance to the use of working areas and amount of angular space used.

6 Limitations

There are some limitations in our study design. First, the datasets we used might not have been large enough – participants didn't feel like they lacked for space in the large display condition. Further studies should use even larger datasets in order to stress the use of space even further. Our findings also suggest that how participants use space impacts their sensemaking performance. Future studies should measure spatial ability of participants to better understand how their management and understanding of space affects performance.

Our participant population may also impact our results. Despite requiring 5-6 hours of work from each participant, an N of 12 is still small and affected our ability to measure significance. Furthermore, novice users may not be the best population to measure the effect of our conditions on sensemaking tasks. Other studies have leveraged professional analysts in order to better understand their user needs, and IST could be similarly evaluated [7].

7 Conclusions & Future Work

IST is a promising approach for sensemaking of large non-quantitative datasets as compared to traditional 2D displays. We found that increasing the amount of space a user has for sensemaking does increase their satisfaction and lower frustration with the environment, as well as creating more varied strategies. The expansive 3D space in IST has the potential to support these strategies better than bounded 2D spaces, despite users' familiarity with the latter approach. The "working area" concept has merit in separating tasks to certain spatial areas, allowing users to leverage their spatial memory in order to better process the data at hand. Similarly, users leveraged more spatial memory to find documents as the available space increased. Furthermore, users that utilized as much space as they were given tended to achieve higher scores on the sensemaking task, though this finding needs to be verified in future studies. Overall, we recommend providing users with as much space as available, which would indicate that we should further explore the use of 3D immersive space for analytical tasks.

Participants had particular issues with IST's user experience, which hampered some results. However, these could be addressed in future work. One pain point in particular, where users had issues switching between document manipulation and text entry, could be solved by using hand tracking instead of 6DOF controllers. This would allow users to move from document organization to text entry and back without putting down or picking up hardware. Since annotation is a common feature for sensemaking, we suggest future immersive analytics applications move away from 6DOF controllers to afford for a better user experience.

Continuing with that theme, future work should focus on a few particular areas. We should create a set of guidelines on best practices and strategies for employing immersive space for the sensemaking task, such as separating work spaces for different specific tasks. Furthermore, HWDs offer a suite of sensors that we can leverage for understanding users' intentions and interests. For example, eye tracking can be explored to understand items of interest to analysts. Lastly, future work could address the storytelling phase of sensemaking. How users leverage immersive space could assist with synthesizing observations and conveying results to others.

Acknowledgments

This work was supported in part by a grant from the Office of Naval Research and NSF grant #CSSI-2003387.

References

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

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

[3] C. Andrews, A. Endert, B. Yost, and C. North. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. *Information Visualization*, 10(4):341–355, 2011.

[4] C. Andrews and C. North. Analyst's workspace: An embodied sensemaking environment for large, high-resolution displays. In *2012 IEEE conference on visual analytics science and technology (VAST)*, pp. 123–131. IEEE, 2012.

[5] C. Andrews and C. North. The impact of physical navigation on spatial organization for sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2207–2216, 2013.

[6] R. Ball, C. North, and D. A. Bowman. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 191–200, 2007.

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

[8] L. Bradel, A. Endert, K. Koch, C. Andrews, and C. North. Large high resolution displays for co-located collaborative sensemaking: Display usage and territoriality. *International Journal of Human-Computer Studies*, 71(11):1078–1088, 2013.

[9] G. Cetin, W. Stuerzlinger, and J. Dill. Visual analytics on large displays: exploring user spatialization and how size and resolution affect task performance. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pp. 1–10. IEEE, 2018.

[10] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, et al. Immersive analytics. in 2015 big data visual analytics (bdva). In *IEEE, Sept*, pp. 1–8, 2015.

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

[12] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. G. Robertson, and G. K. Starkweather. Toward characterizing the productivity benefits of very large displays. In *Interact*, vol. 3, pp. 9–16, 2003.

[13] K. Davidson, L. Lisle, K. Whitley, D. A. Bowman, and C. North. Exploring the evolution of sensemaking strategies in immersive space to think. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

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

[15] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne, et al. Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2021.

[16] B. Ens, S. Goodwin, A. Prouzeau, F. Anderson, F. Y. Wang, S. Gratzl, Z. Lucarelli, B. Moyle, J. Smiley, and T. 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.

[17] B. M. Ens, R. Finnegan, and P. P. Irani. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3171–3180, 2014.

[18] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[19] C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier, and R. Vuillemot. Fiberclay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE transactions on visualization and computer graphics*, 25(1):704–714, 2018.

[20] Y. Jansen, J. Schjerlund, and K. Hornbæk. Effects of locomotion and visual overview on spatial memory when interacting with wall displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[21] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE intelligent systems*, 21(4):70–73, 2006.

[22] W. Lages, Y. Li, L. Lisle, T. Höllerer, and D. Bowman. Enhanced geometric techniques for point marking in model-free augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 301–309. IEEE, 2019.

[23] B. Laugwitz, T. Held, and M. Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pp. 63–76. Springer, 2008.

[24] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[25] L. Lisle, X. Chen, E. J. K. Gitre, C. North, and D. A. Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE 6th Workshop on Everyday Virtual Reality (WEVR)*. IEEE, 2020.

[26] L. Lisle, K. Davidson, E. J. Gitre, C. North, and D. A. Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 529–537. IEEE, 2021.

[27] J. Liu, A. Prouzeau, B. Ens, and T. Dwyer. Effects of display layout on spatial memory for immersive environments. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):468–488, 2022.

[28] W. Luo, A. Lehmann, H. Widengren, and R. Dachselt. Where should we put it? layout and placement strategies of documents in augmented reality for collaborative sensemaking. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2022.

[29] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics*, 20(12):1633–1642, 2014.

[30] J. Mann, N. Polys, R. Diana, M. Ananth, B. Herald, and S. Platel. Virginia tech's study hall: A virtual method of loci mnemotechnic study using a neurologically-based, mechanism-driven, approach to immersive learning research. In *2017 IEEE Virtual Reality (VR)*, pp. 383–384. IEEE, 2017.

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

[32] M. Mcgill, A. Kehoe, E. Freeman, and S. Brewster. Expanding the bounds of seated virtual workspaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(3):1–40, 2020.

[33] H. Park and S. Kim. Do augmented and virtual reality technologies increase consumers' purchase intentions? the role of cognitive elaboration and shopping goals. *Clothing and Textiles Research Journal*, p. 0887302X21994287, 2021.

[34] L. Pavanatto, C. North, D. A. Bowman, C. Badea, and R. Stoakley. Do we still need physical monitors? an evaluation of the usability of ar virtual monitors for productivity work. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 759–767. IEEE, 2021.

[35] J. Ping, Y. Liu, and D. Weng. Comparison in depth perception between virtual reality and augmented reality systems. In *2019 ieee conference on virtual reality and 3d user interfaces (vr)*, pp. 1124–1125. IEEE, 2019.

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

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

[38] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. Van Dantzich. Data mountain: using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 153–162, 1998.

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

[40] D. R. Schwandt. When managers become philosophers: Integrating learning with sensemaking. *Academy of Management Learning & Education*, 4(2):176–192, 2005.

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

[42] I. A. Tahmid, L. Lisle, K. Davidson, C. North, and D. 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)*. IEEE, 2022.

[43] D. S. Tan, R. Pausch, J. K. Stefanucci, and D. R. Proffitt. Kinesthetic cues aid spatial memory. In *CHI'02 extended abstracts on Human factors in computing systems*, pp. 806–807, 2002.

[44] Varjo. Varjo xr-3 technical specifications, https://varjo.com/products/xr-3/, 2023. Accessed on June 12, 2023.

[45] A. Voit, S. Mayer, V. Schwind, and N. Henze. Online, vr, ar, lab, and in-situ: Comparison of research methods to evaluate smart artifacts. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–12, 2019.

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

[47] H. Wu, M. Mampaey, N. Tatti, J. Vreeken, M. S. Hossain, and N. Ramakrishnan. Where do i start? algorithmic strategies to guide intelligence analysts. In *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics*, pp. 1–8, 2012.

[48] L. Zhang, D. A. Bowman, and C. N. Jones. Exploring effects of interactivity on learning with interactive storytelling in immersive virtual reality. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. IEEE, 2019.

[49] Y. Zhang, B. Ens, K. A. Satriadi, Y. Yang, and S. Goodwin. Defining embodied provenance for immersive sensemaking. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2023.

[50] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE transactions on visualization and computer graphics*, 24(1):340–350, 2017.