Exploring the Impact of Augmented Reality on Collaborative Decision-Making in Small Teams

by

Michael Kintscher

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2020 by the
Graduate Supervisory Committee:

Chris Bryan, Co-Chair
Ashish Amresh, Co-Chair
Dianne Hansford
Erik Johnston

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

While significant qualitative, user study-focused research has been done on augmented reality, relatively few studies have been conducted on multiple, co-located synchronously collaborating users in augmented reality. Recognizing the need for more collaborative user studies in augmented reality and the value such studies present, a user study is conducted of collaborative decision-making in augmented reality to investigate the following research question: "Does presenting data visualizations in augmented reality influence the collaborative decision-making behaviors of a team?" This user study evaluates how viewing data visualizations with augmented reality headsets impacts collaboration in small teams compared to viewing together on a single 2D desktop monitor as a baseline. Teams of two participants performed closed and open-ended evaluation tasks to collaboratively analyze data visualized in both augmented reality and on a desktop monitor. Multiple means of collecting and analyzing data were employed to develop a well-rounded context for results and conclusions, including software logging of participant interactions, qualitative analysis of video recordings of participant sessions, and pre- and post-study participant questionnaires. The results indicate that augmented reality doesn't significantly change the quantity of team member communication but does impact the means and strategies participants use to collaborate.

I dedicate this work to the many professors who helped me find my interests, the LGBTQ+ SunDevil community who helped me find myself, and Arizona State University for providing a space for it all. A special thank you to my friends and family whose support got me this far, and to my thesis committee members for the countless hours, thoughtful discussions, and genuine interest and acceptance that made this possible.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter 1

INTRODUCTION

This thesis describes a user study that was conducted to explore the impact presenting data visualizations in augmented reality has on user communication and collaboration behaviors when pairs of users work together to analyze visualized data. Users performed closed and open-ended tasks to evaluate attributes of and trends in the presented data, and repeated these tasks on a desktop monitor for a baseline. Rick contextual data was collected and analyzed through software logging of participant interactions, qualitative analysis of video recordings of participant sessions, and pre- and post-study participant questionnaires.

1.1 Motivation for Study

This study was inspired by the Decision Theater at Arizona State University [71]. Decision Theater is a "Decision-Visualization Environment" (DVE) [51] and has been described as "*a multi-million-dollar semi-immersive virtual environment facility purported to help policy makers and the larger community in making decisions about scientific issues and in visualizing output of predictive and scenario-based models*" [29].

The stakeholders of several DVE's have expressed a desire to develop an offering of mixed reality experiences as a modality for participants [51], including Decision Theater.

While this desire exists, in discussing the results of their pilot study on the effectiveness of Decision Theater's methodology, Edsall and Larson caution against confusing a positive response from participants to the "wow" factor of a novel environment for data visualization with a more genuine increase in content understanding or effectiveness of decision making among those participants [29]. Additionally, Isenberg et al. highlight in their survey of information visualization research a need to encourage research on collaborative data visualization, saying that it is critical to more informed decision making [46]. Lam et al. note in their own survey that evaluations of collaborative information visualization systems remain relatively rare [63].

The field of augmented reality has also had relatively few interaction-focused user studies. Swann et al. published one of the first large scale surveys of augmented reality user studies in 2004, calling for a transition from studies focusing on engineering challenges of augmented reality to studies focusing on user interactions with the medium (augmented reality) itself [86]. In a survey of augmented reality evaluation papers between 1993 and 2007, Dünser et al. found collaboration to be the least studied of the categories of augmented reality user studies [28] proposed by Swann et al. In another survey of augmented reality usability studies published between 2005 and 2014, Dey et al. concluded that collaboration remained one of the least-studied applications for augmented reality, and noted an opportunity for more of such studies [25]. Despite the relative lack of studies, Gabbard and Swan argued that user-based studies are critical for the development of emerging technology, including augmented reality [34]. This suggests a significant opportunity exists in performing more user studies on collaboration in augmented reality. Such studies establish sets of use cases where augmented reality is and is not beneficial in supporting collaborative work, and provide insight into how visualization techniques and supporting software can be improved to better support such collaboration.

1.2 Research Question

Recognizing the need for more collaborative user studies in augmented reality and the value such studies present, we propose a user study of collaborative decision-making in augmented reality to investigate the following research question: "*Does presenting data visualizations in augmented reality influence the collaborative decision-making behaviors of a team?*" In much the same way that Tang et al. argued "*assembly task[s] combined the essential elements of AR computer assistance*" [90], data visualization is the considered application in this study because it combines the essential elements of collaboration: a problem to solve that has hidden layers of meaning for interpretation and is complex enough that multiple perspectives can be beneficial. This study will thus explore this research question by presenting several common data visualizations used in DVE's to participants in augmented reality and on a desktop computer and observing participants as they collaborate in teams of two to perform common data analysis tasks on those visualizations.

Chapter 2

BACKGROUND AND RELATED WORK

The user study presented in this thesis draws from three significant areas of research: virtual and augmented reality, information visualization, and collaboration. Virtual reality presents a computer-mediated version of reality that is entirely "virtual" - not physically present in reality - but in doing so "*deprives us of the ability to see the actual world in which we live*" [52]. Augmented reality "*combines real and virtual objects in a real environment, is interactive in real time, and registers and aligns real and virtual objects with each other*" [21]. Information visualization is the practice of representing information in a graphical way to take advantage of the high relative processing power of the human eye [73]. Collaboration is when multiple people are working together towards a single task or goal [37]. In this study, augmented reality is used as the modality to present information visualizations, which are collaboratively analyzed by pairs of users.

2.1 Virtual and Augmented Reality

Virtual reality and augmented reality both fall on the "mixed reality" spectrum, which is defined based on the mix of physical reality (the real world a person experiences with their own senses and without any computer assistance, or "mediation") and virtual reality (the world a computer creates for a user's senses, but does not physically exist) [21]. As suggested, virtual reality falls on one end of this spectrum, where the computer is creating everything the user sees, and the physical world around the user is completely blocked out. Figure 2.1 shows the mixed reality spectrum. As shown in Figure 2.2, augmented reality falls somewhere in the middle of this spectrum

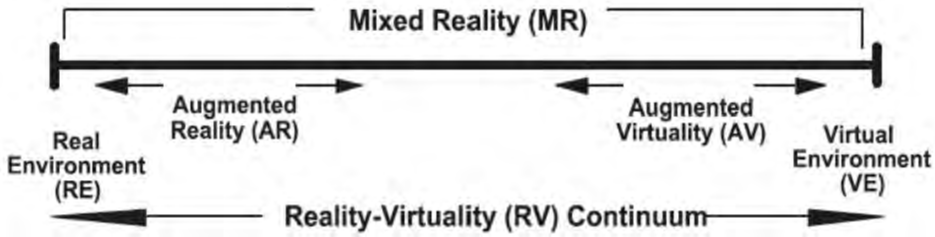

Figure 2.1: The Mixed Reality Spectrum, as Depicted by Schnabel et al. [83]

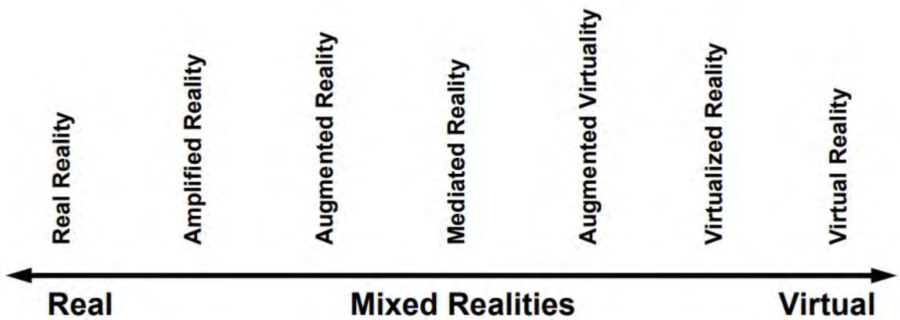

Figure 2.2: Reality Concepts on the Mixed Reality Spectrum, as Depicted by Schnabel et al. [83]

because it creates some aspects of the world the user experiences, while also allowing most or all of the physical world to still be observed. Augmented reality blends these two aspects together into a single, "mixed" reality perceived by the user. Table 2.1 summarizes some of the differences between augmented reality and virtual reality.

Figure 2.3: A Depiction of Mixed Reality Devices from Microsoft [15]

Augmented Reality (AR)	Virtual Reality (VR)
Places virtual objects in the physical world	Both objects and world are virtual
Interactive in real-time	Interactive in real-time
Viewport driven by the user moving the device	Viewport driven by the user moving the device
Does not mediate verbal communication	Does not mediate verbal communication
May augment, or add to, but does not mediate nonverbal communication	Must completely represent, and thus mediate nonverbal communication

Table 2.1: Comparing AR and VR

2.1.1 Value of Augmented Reality

Why use augmented reality in collaborative work? Since the field of Computer-Supported Cooperative Work (CSCW) was first defined more than three decades ago [36], many studies have been performed to evaluate what augmented reality can contribute towards better assisting people in working together. Several of these of these studies have found augmented reality contributes to collaboration [3, 7, 18, 27, 90]

Schmalstieg et al. pioneered an early collaborative augmented reality system in the 1990's they named "Studierstube", in which multiple users see the same virtual content overlaid on the physical world in front of them [87, 82]. They identified five key advantages to collaborative mixed reality [7]:

- Virtuality: Objects that don't exist in the real world can be viewed and examined.
- Augmentation: Real objects can be augmented by virtual annotations.
- Cooperation: Multiple users can see each other and cooperate in a natural way.

- Independence: Each user controls their own independent viewpoint.
- Individuality: Displayed data can be different for each viewer

Previous user studies have demonstrated augmented reality is useful for collaboration on assembly tasks [1, 3, 12, 74, 90]. Baird and Barfield showed that participants preferred augmented reality instructions to instructions printed on paper or displayed on computer monitors when working on manual assembly tasks [3]. Tang et al. performed a similar user study, asserting that "*assembly task[s] combined the essential elements of AR computer assistance: (1) Spatial registration of virtual and real objects, (2) Interaction of virtual and real objects, and (3) The use of AR to sequence and coordinate human procedural action*". They observed decreased task completion times and significantly decreased error rates among participants who used augmented reality as opposed to other means of instruction [90]. Nee et al. provide a thorough overview of the research on augmented reality in manufacturing, which assembly tasks are a part of [74].

Augmented reality has also shown strong potential for bridging gaps across many different fields. Dong et al. discussed the value that augmented reality can provide to diverse engineering teams and engineering students alike [27]. Broll et al. discussed the value that augmented reality brings to decision makers with diverse expertise/backgrounds. They observed that projects that demand highly cooperative tasks, such as those in the architectural and urban planning fields, frequently follow a pattern of individual work with meetings only to discuss the individual work. They proposed that augmented reality can bring collaboration between team members with differing expertise into the production of shared project deliverables, such as sketches and design documents [18].

2.1.2 Types of Augmented Reality Devices

Augmented reality displays can be divided into three main categories, based on how the user wears or otherwise interacts with the display: handheld displays (HDDs), head-mounted displays (HMDs), and projector displays (sometimes referred to as "Spatial Augmented Reality" or SAR) [10] [74].

Handheld Displays (HHDs) are devices that a user holds in their hand while operating. The device has a camera on the back (pointed away from the user) and captures a live video feed. The device then blends the captured feed with the virtual images and displays the mixed image on its screen (facing the user). Tablets and mobile phones have become commonplace in this manner as augmented reality devices, and are sometimes referred to as Mobile Augmented Reality (MAR) devices [21].

Head-mounted displays (HMDs) are devices that a user wears on their head, and come in two different types: optical see-through (OST) and video see-through (VST) [74, 77]. OST-HMDs are head mounted displays that have a see-through component to them the user can look through to directly see the real world in front of them. Virtual images are then projected onto the see-through component so that they are superimposed onto the real world from the perspective of the user. Figure 2.4 shows the Microsoft HoloLens 2 as an example of an OST-HMD. The see-through eye-covering can be seen on the front of the visor (to the left in the image). VST-HMDs are head mounted displays that completely enclose the user's eyes and obscure the user's view of the natural world, like a virtual reality (VR) headset would. To enable the user to see, cameras are attached to the headset and a live video feed from the cameras is played on screens in the headset in front of the user's eyes. Virtual images are then mixed with the captured footage from the camera and presented to the user as a part of the live video feed. Figure 2.5 shows the Samsung HMD Odyssey+ as an

Figure 2.4: OST-HMD Example: HoloLens 2 [15]

Figure 2.5: VST-HMD Example: Samsung HMD Odyssey+ [15]

example of a VST-HMD. Two front-mounted cameras can be seen on the visor part of the headset.

Projector displays, also referred to as Spatial Augmented Reality (SAR), use mirrors, screens, and projectors to project or display augmentations onto physical objects [10].

2.1.3 Perception in Augmented Reality

Significant work has been conducted on perception in augmented reality. Kruijff et al. summarized and classified these issues into five categories: Environment, Capturing, Augmentation, Display Device, and User [61]. From the issues listed in these five categories, only the subset of issues discussed here were considered in developing the user study for this thesis: Stereoscopy, Field of View, Visual Distortion, and Occlusion. Table 2.2 summarizes how these issues affect the three types of augmented reality devices described in section 2.1.2 and virtual reality devices.

Issue	Augmented Reality (AR)				Virtual Reality (VR)
	OST-HMD	VST-HMD	HHD	SAR	
Stereoscopy	Inherently supports	Can support	Limited in ability to support	Can support	Can support
Field of View	Limited	Limited	Very Limited	Not Limited	Not Limited
Visual Distortion	Some possible from see-through material	Some possible from camera placement and camera lens effects	Some possible from camera placement and camera lens effects	Some possible from material projected onto	N/A
Occlusion	Some	Some	Minimal	None	Complete

Table 2.2: Summary of Perceptual Issues Impacting AR and VR

Stereoscopy

A stereoscopic view has a slightly different viewpoint presented to each of the user's eyes, and thus creates true three-dimensional vision, as the user's eyes naturally do. HHD's are limited in their ability to present a stereoscopic view. Because mobile devices generally have a single rear-facing camera, they are limited to a monoscopic view and do not support a truly stereoscopic view [8] [9]. HMDs can support stereoscopic views, as can SAR [10].

Although Dohse concluded that stereoscopy does not appear to have a significant effect on the memory of the users, he noted it may have an indirect effect by reducing frustration, which is strongly correlated with the user's memory [26]. Belcher et al.

concluded that stereoscopic viewing had little impact on performance when analyzing 3D graphs [5], but Billinghurst et al. reasoned that a lack of stereoscopic viewing may have negatively impacted the AR cases in their study [9]. These mixed results suggest that stereoscopy, or a lack thereof, is only a factor in certain conditions rather than a blanket benefit or drawback of augmented reality in general.

Field of View

A field of view measures how much of the mixed reality world the user can view through the device at once [61]. The impact of a limited field of view is well known to increase the difficulty of a wide range of tasks [93]. For example, Ishikawa et al. proposed that the smaller viewing area of a device was a cause for the degradation in navigation-based tasks they observed with their users [48]. There are plenty of studies, however, that have found aspects of user performance a limited field of view does not impact. According to Dohse, an increased field of view appears to have no impact on the user's memory [26]. Knapp and Loomis concluded that a limited field of view also had no impact in an AR user's ability to perceive distance, suggesting that some other issue was the cause of frequently reported distance perception issues with augmented reality devices [56]. Kishishita et al. conducted a user study that explored the effects on users of increasing the field of view in an HMD, and concluded that field of view had little impact on user response time and mental load [54].

HMD's and HHD's both suffer from a limited field of view, whereas SAR's do not [10]. Several systems have been proposed to address this issue in HMD's, however. Baudisch and Rosenholtz proposed a visualization technique they called "Halo" to address the problem of drawing user's attention to off-screen objects on 2D maps. Halo involved drawing a circle large enough around the object so that part of the edge of the circle (an arc) would be on-screen. The user could then gauge how far away the off-screen object was based on how tight the visible arc was [4]. Renner and Pfeiffer built upon this to propose "Spherical Wave-Based Guidance" (SWAVE) as an adaptation of the technique for the 3D environment inherent to augmented reality. They tested SWAVE in a user study against several arrow-based techniques, and found participants favored 3D arrows, with SWAVE being a strong alternative [79].

Visual Distortion

Visual distortion is any unintended alteration of the physical environment the device creates for the user [61]. VST-HMDs suffer from what is known as "visual displacement", whereas OST-HMDs do not [11]. This is a type of sensory rearrangement caused by the difference in position between the user's eyes and the camera lenses. The cameras record from a slightly different perspective relative the world and the user's body, thus causing some disorientation. This leads to a decrease in accuracy when the user attempts to take on certain tasks [11]. Research suggests there are ways to overcome this, such as the suggestions proposed by Biocca and Rolland [11]. They suggest reducing the distance between the camera and eye and tiling screens to compensate for the changed camera hardware size/position. While some visual displacement is unavoidable (as zero would mean the cameras and eyes are in the same physical position), they also note that pointing errors arising from the visual displacement may become negligible before the visual displacement reaches zero [11].

OST-HMDs do still have some visual distortion, however. Such distortion is caused by the incoming light from the real world being refracted through the see-though component of the OST-HMD, similarly to light passing though lenses on regular glasses [49].

Occlusion

Both OST-HMDs and VST-HMDs suffer from some occlusion (blocking) of the user's view of the physical environment [11], whereas SAR's do not [10]. This is due to the presence of the device in the user's field of view. Navab and Qian present some techniques to combat the occlusion caused by OST-HMDs, including putting LEDs around the screen to turn on when an object is detected in the occluded part of the field of view, or indicators on the screen to likewise point towards such objects [77].

2.1.4 Comparing Augmented and Virtual Reality

Krichenbauer et al. discuss the difficulties in directly comparing augmented reality and virtual reality, including the difference in interaction with physical objects, and determining how to represent the environment in virtual reality, along with discussing several previous studies that struggled with this task. They then ran a user study to directly compare user performance in augmented reality and virtual reality, and concluded that users performed better in augmented reality [60].

McGill et al. noted that one issue that impacts virtual reality users more than augmented reality users is a sense of isolation from their collaborators, and then proposed addressing this isolation by adding real-world elements into the user's view [70]. This change moves the user's experience further from the "virtual reality" end of the mixed reality spectrum into the "augmented reality" end [21]; the proposed solution to this issue in virtual reality is to make the experience more like augmented reality.

Another key difference between augmented and virtual reality is the level of mediation. The user's awareness of the mediation is referred to by Lobard et al. as "transparency" [67]. According to Marsh, the level of presence depends on the transparency of an interface [69]. Tang et al. used the ITC-SOPI questionnaire developed by Lessiter et. al. [64] to compare the sense of presence between augmented and virtual reality users. They found augmented reality to have a higher sense of presence, and attributed this to the lower level of mediation augmented reality requires compared to virtual reality [89].

The visibility of collaborator's hands and gestures represent an aspect of presence that is important to the effectiveness of a collaboration system. Kirk and Fraser conducted a user study comparing different representations of a remote collaborator's gestures, and concluded that unmediated video of the collaborator's hands is the best system design recommendation - no virtual representation they tested worked as well [53]. Alem and Li found a contradictory result in their own similar user study where participants performed the same across different gesture representations. Their results also showed, however, that directly showing a remote collaborator's hand gestures increases the user's experience and perception of collaboration quality [1]. These results demonstrate that representing users' hands to fellow collaborators can improve user performance and experience. Augmented reality systems can inherently show a user's hands, whereas virtual reality systems must have representations of the user's hands created for the experience.

There are mixed research findings on how an OST-HMD may impact a user's accuracy in locating, selecting, and communicating the selection of virtual objects. Büschel et al. found no significant difference in error rates between spatial and touch modalities [19]. Participant performance was similarly consistent across augmented reality and virtual reality modalities in Müller et al. [72].

2.2 Information Visualization

Information Visualization (InfoVis) is the practice of presenting data (information) in a visual format (charts, or data visualizations) for a human to analyze. Information visualization exploits the large cognitive data processing capacity of the human visual system, and as such has significant advantages over other techniques for presenting data [68, 73]. Isenberg et al. describe the focus of information visualization as being a tool to inform decision-making, and call this a differentiating factor from other computer supported collaborative work (CSCW). They also make a call for integrating collaboration support into information visualization systems from the very beginning, reiterating the essential role information visualization plays in quality collaborative decision-making [46]. Bresciani and Eppler demonstrated that using visualizations in collaborative work environments can improve outcomes of knowledge sharing, even if the collaborating workers weren't consciously aware of the improvement [17]. As an example of how this is valuable to collaborating decision makers, Rohrer discussed the advantages of visualizing information as a means to increase the accessibility of data to decision makers in engineering processes [81]. 3D information visualization is also particularly attractive to decision makers. Raja et al. found higher levels of immersion in data visualizations may aid users in evaluating the presented data [78]. Liu et al. provide a recent comprehensive survey of research in information visualization [66], but the two major aspects of information visualization of interest to this paper are graphical perception and tasks.

2.2.1 Graphical Perception

Perception is concerned with how a user reads the data presented in the charts. Data attributes are visually encoded and arranged in charts to be presented to a user. Munzner describes marks as the basic geometric primitives that encode a data point, and channels as the visual grouping of marks that encode an attribute of the data. Examples of marks include points, lines, arcs, and areas. Examples of channels include spatial position, color, area, and size [73]. Users read the data encoded in the chart via the process of visual decoding, also referred to as graphical perception [23].

Cleveland and McGill established an ordering of channels based on perceptual accuracy - how easily a user could read the correct information from that channel - in their seminal work on graphical perception. They found spatial channels to be the most accurate [24]. This has been corroborated through various studies since. Even two-and-a-half decades later, Heer at al. duplicated these results with more modern tools during their evaluation of using Mechanical Turk as a research tool [39]. Wigdor et al. extended Cleveland and McGill's work by examining how distortion caused by different perspectives and viewing angles inherent in multi-user environments can affect graphical perception. They saw some changes in perceptual accuracy based on viewing position, such as favoring length encoding over position encodings for certain positions and avoiding angle and area encodings entirely. They did note, however, that the perceptual differences they observed in their user study were relatively small [97].

Graphical perception in 3D presents a unique set of challenges relative to its 2D counterpart. Munzner describes problems with graphical perception in 3D, and offers that 3D visualizations must be clearly justified over 2D as the rule of thumb [73]. Kyritsis et al. presented a user study examining the usability of 3D interfaces, and concluded that 2D interfaces were more usable [62]. A notable limitation to their study, however, is that they examined interfaces based on the WIMP (Window, Icon, Menu, Pointer) paradigm, and thus may not be applicable to modern augmented reality interfaces for presenting charts. Brath outlines several use cases where 3D interfaces can be effective for information visualization, particularly by taking advantage of the additional spatial channel the third dimension 3D visualizations offer over 2D visualizations. [14]. The work of Wigdor et al. on distortion from perspective [97] is useful in guiding the selection of chart types to display in 3D because, as Brath notes, 3D visualizations intrinsically include such perspective [14]. Chun et al. explored the practice of redundantly encoding data into charts, where a data attribute is encoded onto multiple channels, and found no positive nor negative effect on graphical perception. Notably, their study was limited to value and position channel pairs and only evaluated 2D visualizations viewed head-on [22]. Redundantly encoding data thus may still be beneficial in 3D visualizations as a means to dilute the effect of different perspectives on a user's graphical perception observed by Wigdor et al. [97].

2.2.2 Tasks

Tasks are the actions a user performs when viewing or interacting with visualized data. Munzner describes a taxonomy of abstract tasks common to all data visualizations using pairs of actions and targets. Each task can be seen as an action and a target for that action [73]. Brehmer and Munsner connect this taxonomy to the vocabulary from previous work, and describe how it mirrors the analytical thinking process [16].

Yi et al. classified the interaction techniques common to all information visualization into seven categories based on the user's objective with the interaction: select, explore, reconfigure, encode, abstract/elaborate, filter, and connect [98].

Certain tasks are better supported by 3D data visualizations. Springmeyer et al. observed scientists using 3D visualizations to gain a sense of context in the data they were analyzing, before turning to 2D visualizations to observe detail views [84]. St.

John et al. explain some of these observations by proposing the use of functional distinctions in the data visualization tasks to be performed to determine if a 2D or 3D visualization would be more useful [85]. Based on this proposal, the previously discussed findings from Kyritsis et al. that favored 2D interfaces over 3D ones [62] can be attributed to a mismatch between the tasks the users were trying to carry out and the choice to use a 3D representation of the data instead of a 2D one. Tory et al. reinforce this proposal with findings from their own user study, concluding that certain functional groups of tasks such as those that required orientation, navigation, and viewing a larger context were better suited for 3D visualizations, while other functional groups of tasks such as those requiring precise measurements were better suited for 2D visualizations [91].

2.3 Collaboration

Ellis et al. defined "groupware" as "*computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment*". They also established the three fundamental concepts that computer supported cooperative work (CSCW) must support: Communication, Collaboration, and Coordination [30]. Communication is how users exchange information with each other during cooperative work on a shared task. Collaboration is how users actually jointly interact with the task. Coordination is how users organize each other's work to together complete the task. Communication and Collaboration are of particular interest to this paper, and as Heer et al. describe, a look at CSCW research is relevant to collaborative information visualization [40].

Several classifications of collaborative systems have been proposed. Johansen et al. described a taxonomy for collaborative interaction using a function of space and time. Interactions could be either same or different place, and same or different time, for a total of four categories. These categories are displayed in Table 2.3 [50]. Ellis et al. then applied this taxonomy to CSCW systems [30], and Munzner describes this in the context of collaborative information visualization [73]. Two (or more) users in the same room collaborating at the same time would be considered “synchronous co-located” by this taxonomy. Nee et al. classified AR-based collaborative design systems into two sub-categories [74]. Visualization-based design systems provide an environment where designers can visualized, annotate, and inspect 3D models collaboratively. Co-design systems allow designers to create and modify 3D models collaboratively. Most collaborative AR-based systems are visualization-based [74]. Isenberg et al. note in their survey of collaborative information visualization that such classifications for collaborative systems are useful for scoping research focuses [46].

		Time	
		Same	Different
Space	Same	Same-Time/Same-Space	Different-Time/Same-Space
	Different	Same-Time/Different-Space	Different-Time/Different-Space

Table 2.3: A Taxonomy of Time and Space for Collaborative Groupware Systems

Dey et al. observed in their literature survey of augmented reality papers that most collaboration user studies use within-subject designs, and report both qualitative and quantitative results. Studies were pretty evenly split between using HMDs, HHDs, and alternative display types. Most collaboration studies were also focused on remote collaboration, with Dey et al. noting that face-to-face collaboration was an opportunity for future studies to more thoroughly cover [25]. The user study conducted for this thesis follows the trend of within-subject designs and employing both qualitative and quantitative results, but also contributes to the body of work by helping fill the gap in face-to-face collaboration studies.

2.3.1 Communication

Inkpen observed in their dissertation that grouping children around a single computer screen has motivational and performance benefits when compared to two children using separate computers to collaborate [44]. They reasoned that this was because the computer creates an artificial separation between the task and communication spaces. This was an illustration of a critical concept in evaluating communication between collaborating users in CSCW systems: the overlap of "task space" and "communication space".

Ishii et al. define a "seam" as "*a spatial, temporal, or functional constraint that forces the user to shift among a variety of spaces or modes of operation*". They developed the ClearBoard system in an attempt to remove seams from collaborative software [47]. Billinghurst and Kato establish the definition of a "task space" as "*the shared workspace*" where tasks are performed, and the definition of "communication space" as "*the common interpersonal space*" where collaborating users communicate with one another. They also establish that the task space is often a subset of the communication space in face-to-face conversation. As an example of this, consider two people across the table from each other playing cards together. The task space would be the table with the cards they are both manipulating on it, and the communication space would be the space between them, where they speak towards and look into to see each other's expressions. Applying the work of Ishii et al., they declare a CSCW system that does not present the task space as a subset of the communication space as having a seam when users are forced to switch between the two spaces [7]. When examining communication between collaborating augmented reality users under different space and device configurations, Kiyokawa et al. found OST-HMDs and the positioning of the workspace between the collaborators to require the least amount of verbal communication to accomplish collaborative tasks. OST-HMD's allow users to see each other's nonverbal gestures and cues, and the positioning of the workspace between the users places the task space as a subset of the communication space, as it would be in natural, unmediated conversation [55].

Another critical concept in communication with CSCW systems is "conversational grounding", which is defined as the development of mutual understanding between conversational participants. Kraut et al. present a model for participants collaborating on physical tasks. They establish that visual information is a vital part of collaborative communication because it helps participants gain situational awareness and conversational grounding. Situational awareness is an understanding of the state of the space, and conversational grounding is the shared understanding between participants. [57]. Gergle et al. demonstrated that participants would rely more on verbal communication as their shared view of the task space decreased, and concluded that showing participants what the other is doing is not enough; both participants need a shared understanding of what the other can see [35]. Ou et al. ran a user study tracking the eye gaze of two remote collaborators in a helper-worker relationship who were solving a puzzle. Although the collaborators had no view of each other and only a shared view of the workspace, Ou et al. observed a strong connection between altering conditions that increased the difficulty of communicating about the task and a need for the collaborators to vocally communicate to establish conversational grounding [75].

As a highly visual modality, augmented reality can have a significant impact in shaping conversational grounding between collaborating users. Kraut et al. performed an empirical user study that evaluated the performance and communication of worker/expert pairs collaborating to perform physical repair tasks. They performed the evaluation across several modes of video and audio communication and concluded that experts (those with the knowledge on how to accomplish the task) were more proactive in sharing that knowledge with workers (those actually performing the task) when the expert could see the worker and the worker's efforts [59]. Chastine et al. built upon this by hypothesizing and concluding through their own user study that two collaborating augmented reality users are better able to coordinate actions when sharing a common point of view. They explained this effect in terms of collaborating users being better able to refer to and communicate that they were referring to the same reference points [20]. Müller et al. then hypothesized that co-located collaborators (users who are physically together working on a common task) would benefit more from augmented reality while remote collaborators (users who are not physically together but are still working on a common task) would benefit more from virtual reality. They reasoned that co-located users share the same physical environment and thus could use the physical environment as reference points while communicating, whereas remote users do not share the same physical environment and thus would have to rely on a shared virtual environment to provide the same reference opportunities. Their user study found no significant differences in performance to attribute to this difference, however [72]. A possible limitation of Müller et al.'s study is their use of HHD's. They even noted in their analysis of the user's qualitative feedback that "*polarized opinions towards the AR and VR configuration may not necessarily apply to displays that cover a larger amount of the visual field*" [72]. This would be consistent with the importance of visual information and risk of remote collaboration systems limiting that information described by Kraut et al. [57]. Because HMD's inherently cover more of the user's visual field, and because the study neither encoded nor directly observed collaboration styles nor communication, their results may not be applicable to two co-located, collaborating users working with HMD's.

2.3.2 Collaboration Styles

Tang et al. conducted a pair of user studies with participants engaging in exploratory information-finding on tabletop displays to observe collaborative coupling, the amount collaborators are engaged in each other's work, and proposed a classification of the different collaboration styles they observed. They proposed four styles, ranging between highly engaged and highly disengaged [88]. These styles are illustrated in Table 2.4. Isenberg et al. later extended this work by conducting an observational user study and proposing a slightly different classification of collaboration styles [45]. These styles are illustrated in Figure 2.6. These styles were focused on tabletop displays, and to the best of our knowledge, similar studies that investigate collaboration styles using AR for data visualization and analysis have not yet been conducted.

2.3.3 Decision-Visualization Environments

The concept of using software and hardware to present information to a group of people to support decision-making is several decades old. Huber discusses the concept of such a system, and refers to it as a "Group Decision Support System" (GDSS). Huber notes that a GDSS's usage will depend on how well the system's capabilities match the tasks of the participants, and proposes adopting a "task-driven" strategy of GDSS design, where the tasks of the users should be examined and the system should be designed to support these tasks [43]. Leung discusses the need for a decision support system (DSS) to emphasize the knowledge engineering aspects of using human intelligence in decision making, rather than just the software engineering aspects of such systems. Leung also described a general architecture for DSS's, calling for a DSS development environment which would include a generic framework that could

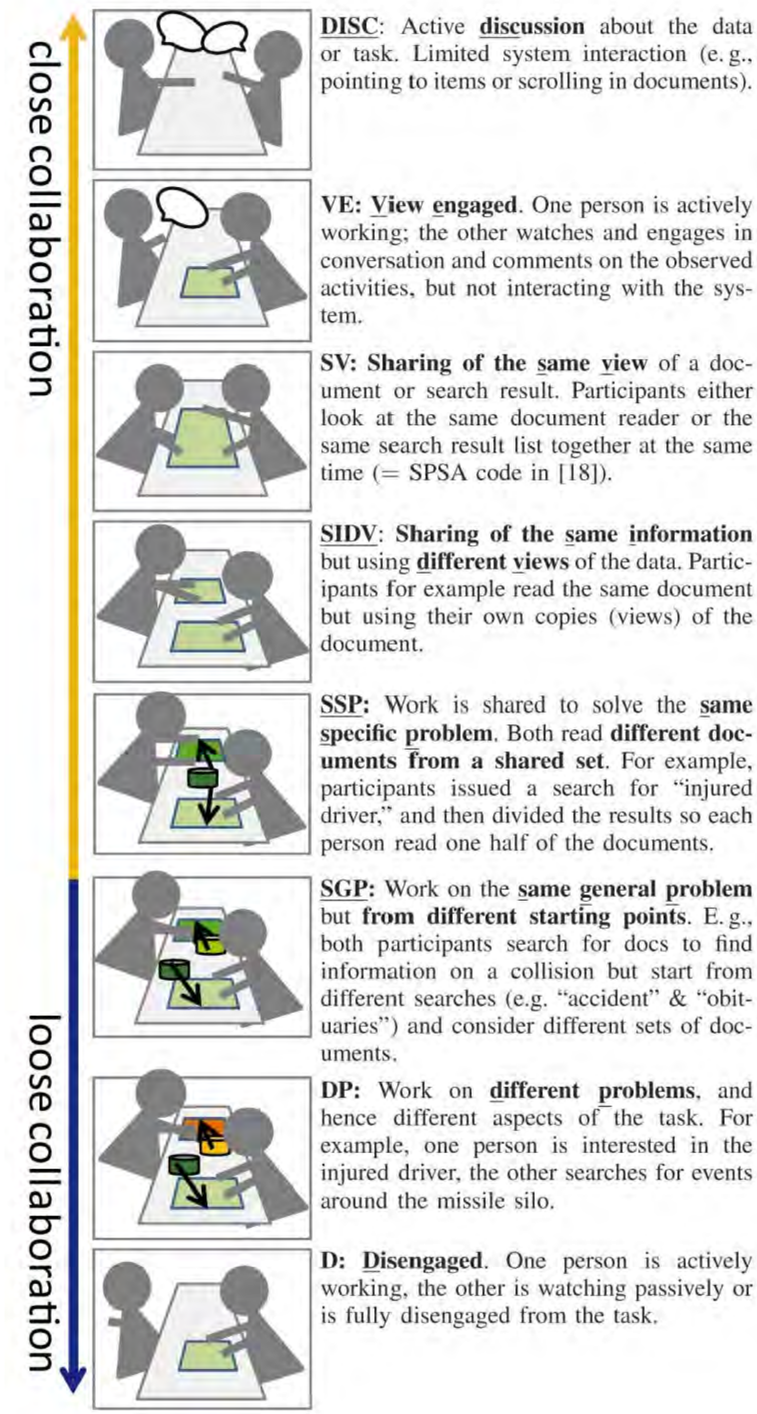

Figure 2.6: Expanded Collaboration Style Classification from Isenberg et al. [45]

Style Name	Description
SPSA: Same Problem Same Area	Collaborators actively work together to evaluate, trace, or draw a route, often accompanied by conversation.
VE: View Engaged	Collaborators work together and often converse, but only one actively manipulates the display while the other watches closely enough to suggest corrections.
SPDA: Same Problem Different Area	Collaborators work simultaneously on the same sub-problem, but focus on different parts of the table, and do not converse. For instance, participants may be evaluating alternate solutions of the same sub-problem.
V: View	One collaborator works on the task, while the other watches but is not sufficiently involved to help or offer suggestions and only reacts to high-level activities, such as when the active person stops working or needs resources (e.g. a widget).
D: Disengaged	One collaborator is completely disengaged from the task, not paying any attention to the task or partner.
DP: Different Problems	Collaborators work completely independently on separate sub-problems at the same time. Each person's interactions with the workspace are not related to the other in any way. Participants often peek at one another to maintain an awareness of the other's activities.

Table 2.4: Collaboration Style Classifications from Tang et al.

be used by domain experts to generate a DSS for their specific problems [65].

As a more modern taken on the same concept, John et al. define a Decision-Visualization Environment (DVE) as a semi-immersive environment that supports research, planning, and decision-making at the science-society-policy interface. They present a nine module framework for abstractly describing the design of DVE's, including one module for visualization. Additionally, they identify three groupings of visualizations used in DVE's, grouped based on the role the visualization plays in the process [51]:

- Summary Graphs: show overall performance or process adaptive throughout an event
- Experiential Visuals: allow exploration of a certain issue and manipulation of scenario parameters
- Focus Visuals: support details of the scenario or issue by delivering additional background information

Boukherroub et al. proposed a generic framework for designing Decision Theaters (a type of DVE) as a part of their work on supporting forest management that emphasized the interaction between "Decision entities" (the people making the decisions) and the system via setting parameters and specifying scenarios [13].

Several valuable and tangible use-cases for DVE's that require collaboration between different stakeholders have already been explored. Hahn et al. provide some evidence that DVE's can better communicate some aspects of ethically complex policy decisions than more traditional presentation mediums, but also warn that such gains may vary depending on the demographics of the participants [38]. Bettencoiurt et al. noted that DVE's are useful for their "*power and flexibility*" in "*computational and informational infrastructure*" as tools in solving challenging epidemiological problems [6]. Studies have also been done exploring examples where the DVE contributed to educating public administration officials [42] and to evaluating uncertainty in decision support systems [95].

Chapter 3

STUDY DESIGN

3.1 Hypotheses

The research question posed for this study is: "*Does presenting data visualizations in augmented reality influence the collaborative decision-making capabilities of a team?*". To investigate this, several hypotheses are proposed for the user study to test.

Büschel et al. observed a nearly even split between the number of participants who moved around significantly and the number of participants who stayed mostly still when carrying out tasks on a data visualization in virtual reality. This behavior persisted regardless of the task presented to the users [19]. Following this, we hypothesize that

H1: *The distribution of quantity of participant movements will not change significantly between tasks.*

An important limitation to note, however, is that Büschel et al. only examined mobile devices, and did not consider HMDs nor desktop displays [19]. Because HMDs afford users inherently more natural movement than mobile devices or desktops, we hypothesize that

H2: *The average number of participant movements will be higher when completing tasks with augmented reality than on a desktop computer.*

Participants directly express that physical objects help them orient themselves when viewing virtual reality content with tablets [19]. Because augmented reality is visually embedded in the participants' environment, surrounded by physical objects that can provide context for communication whereas the chart background of a desktop computer screen provides fewer physical context clues, We hypothesize that

H3: *The amount of gestures participants use when communicating will change significantly between when using the augmented reality headsets and when using the desktop computer.*

Kraut et al. observed collaborator behavior patterns changing depending on how much visual information was available to them [58]. In a later experiment, Gergle et al. further demonstrated that collaborators relied more on verbal communication the less of the work-space they visually shared [35]. Kiyokawa et al. found OST-HMDs to require the least amount of verbal communication between participants to accomplish collaborative tasks. This was because subjects resorted more to speech the more difficult it was to read nonverbal communication cues. Additionally, Kiyokawa et al. observed participants communicated more with gestures than verbal means when the collaborative visualization was placed between the users than to one side of both [55]. It is worth noting that Billinghurst et al. found the contradictory effect of users verbally communicating more and using fewer gestures in their experiments comparing different modalities [9]. These experiments are much older and use older technology than what is available today, however, and are likely subject to the detrimental effects of limited user knowledge due to being an emerging technology noted by Bach and Scapin [2]. Billinghurst et al. even acknowledge that the results of their study were likely impacted by limited field of views and the use of biocular instead of stereoscopic views in the HMDs they tested with [9] – two negative effects more modern HMDs have improved upon. We thus take the more recent results as more accurate, and further hypothesize that

H4: *Participants will verbally communicate less when using the augmented reality headsets than when using the desktop computer.*

Additionally, Fussell et al. found that collaborating users focus their gaze equally as much on their collaborating partner as they do on the tools and task [33]. This suggests that overlapping the communication and task space will be beneficial to users. We thus hypothesize that

H5: *Participants will spend more time positioned such that their communication and task spaces overlap than not overlapping when using augmented reality.*

Isenberg et al. described eight different collaboration styles [45], building upon the work of Tang et al. [88]. We have found little work investigating how tasks and modalities will impact collaboration styles, and thus we take the null hypothesis for each and hypothesize that

H6: *There will be no significant change in collaboration styles between augmented reality and desktop modalities.*

H7: *There will be no significant change in collaboration styles between tasks.*

Büschel et al. observed participants taking significantly longer to complete "navigation tasks", tasks that required finding and navigating to specific points in a visualization, in the spatial interaction modality compared to the touch modality. [19]. All of our tasks include navigation tasks as described. Additionally, Billinghurst et al. observed participants taking longer to solve problems in augmented reality than projected modalities [9]. We thus hypothesize that

H8: *Participants will take significantly longer to complete tasks in the augmented reality modality than in the desktop display modality.*

3.2 Experiment Design

A within-subject user study was conducted to test the given hypotheses. This study consisted of three independent variables: Modality, Chart Type, and Task Type. Two modalities, three chart types, and two task types were tested. Participant

Figure 3.1: Desktop Setup Used in the Study

teams engaged in two trials of each task type on each chart type on each modality, for a total of 24 trials per team ($2 \times 3 \times 2 \times 2 = 24$). Six teams completed the study, resulting in 144 trials. Due to a computer hardware failure, 7 desktop trials for one team were discarded, meaning that a total of 137 trials were collected for analysis and validation of **H1**–**H8**.

Seven dependent variables were monitored during the study: task response time, number of gestures for each participant, number of utterances for each participant, the number of times one of the participants looked at the other, the number of times participants clearly made an effort to establish conversational grounding, which participant controlled the mouse during the desktop trials, and the distance participants moved the view port. These variables were monitored via a combination of visualization software logging and audio/video recording participants.

3.2.1 Modalities

Two modalities have been selected to test across: desktop (D), and augmented reality headset (AR).

Figure 3.2: HoloLens Setup Used in the Study

Desktop (D) is the viewing condition where participants are viewing and interacting with a virtual chart on a 2D monitor. This will serve as the baseline to compare the augmented reality conditions against. Such comparison is common in augmented reality user studies. Participants are able to navigate around the displayed charts by using the mouse to manipulate the view on screen. Holding the right mouse button while moving the mouse rotated the view in the direction the mouse was moved. Holding the middle mouse button (scroll-wheel) while moving the mouse panned the view in the direction the mouse was moved. Scrolling the scroll-wheel moved the view forwards and backwards in the direction the view was facing (analogous to zoom). These controls are similar to the controls for Unity Game Engine, and some other industry 3D modeling software. Basing the mouse control on 3D modeling software used in industry has been used previously to keep study results "*closely related to real-world applications*" [60].

Augmented reality (AR) is the viewing condition where participants are viewing and interacting with a virtual chart embedded in the real world. This arrangement in augmented reality has been used in several user studies, including [72], [60], [55], [20], and [5].

3.2.2 Charts

Chart Types and Presentation Three chart types have been selected to test across: bar charts, scatter plots, and network diagrams. The inclusion of multiple chart types was an intentional design choice, to increase the generality of the results. Additionally, these chart types cover two of the three main groupings for functions of visualizations in Decision-Visualization Environments (DVE's) described by John et al. (the third grouping can include the charts in this study as well, but is characterized by interactivity and thus technically excluded from this study) [51].

Bar charts visualize multidimensional data by encoding a categorical data attribute or index along one axis and encoding a numerical data attribute along another axis. In three dimensions, two categorical data attributes are encoded, each along a different horizontal axis. A data point is represented by a bar and is then placed at the 2D coordinates represented by the two categorical attribute values the data point has. The height of the bar encodes the numerical data attribute for that data point. Both Büschel et al. [19] and Whitlock et al. [96] provide examples of bar charts being used for 3D data visualizations; the former serves as the example for the bar charts used in this study. A bounding box for the bar chart was additionally displayed, as were tick marks on each bar along a standard vertical interval to assist participants in judging relative bar heights. Each axis on the bounding box was colored, with red representing X, green representing Y, and blue representing Z, to assist participants in referencing directions. Figure 3.3 and Figure 3.4 show an example how the bar charts looked on the desktop computer and on the HoloLens 2 respectively.

Scatter plots visualize multidimensional data by encoding data attributes onto graphical axes (spatial channels) in Cartesian space. Each data instance is rendered as a point, and may have additional attributes visualized through value channels on

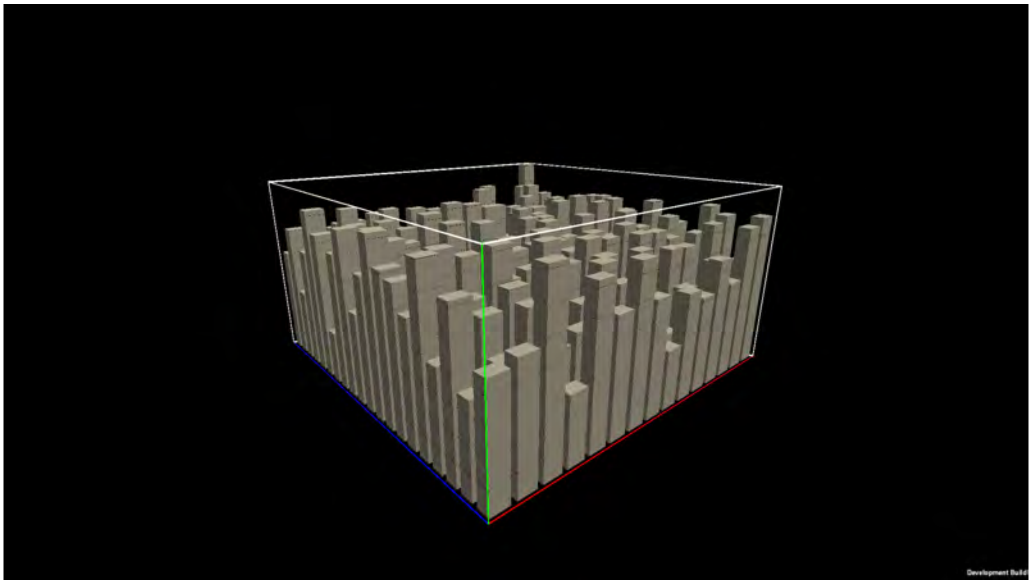

Figure 3.3: Bar Chart as Displayed on the Desktop Computer

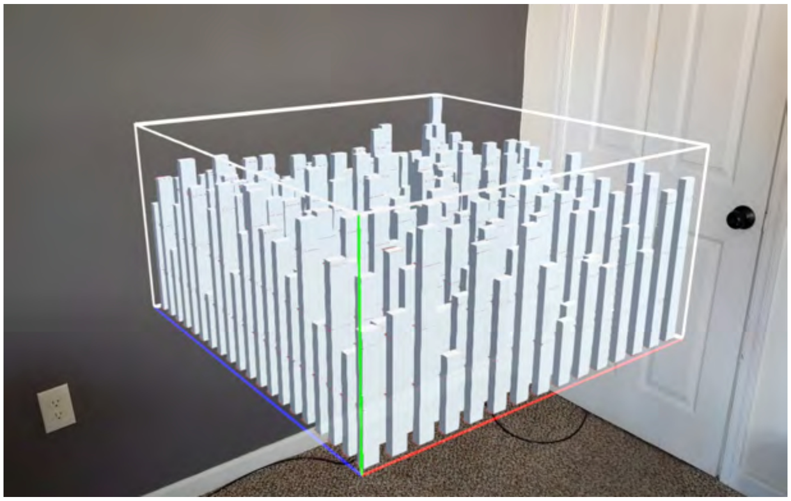

Figure 3.4: Bar Chart as Displayed on the HoloLens 2 AR Headset

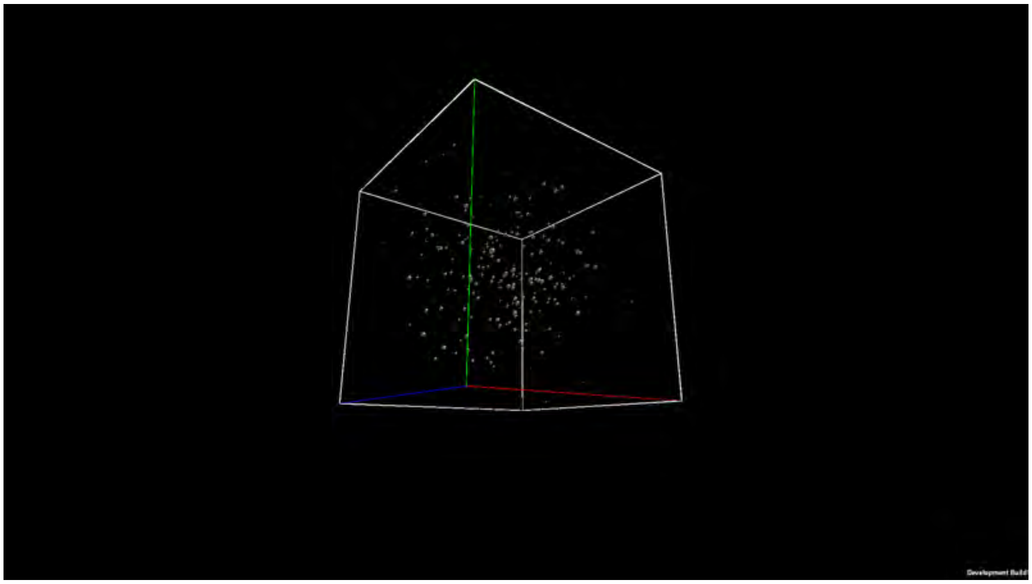

Figure 3.5: Scatterplot as Displayed on the Desktop Computer

that point, such as shape or color [31]. Examples of scatter plots being used for 3D data visualizations include [78], [31], [19], and [96]. The scatter plots used in this study encoded data in the size of the point in addition to the three graphical axis. Points were completely opaque, and thus occluded anything behind them relative to the viewer. A bounding box around the scatter plot was also included to help participants judge relative positions of the points. Each axis on the bounding box was colored, with red representing X, green representing Y, and blue representing Z, to assist participants in referencing direction. Figure 3.5 and Figure 3.6 show an example how the scatterplots looked on the desktop computer and on the HoloLens 2 respectively.

Network diagrams encode relationships or connections between data points, which are arranged in space via a desired layout method. The spatial arrangement of the data points does not encode any information, and thus they are arranged for the user's convenience. Relationships between data points, or "nodes", are encoded by connected related data points with lines, or "edges". Belcher et al. provide an

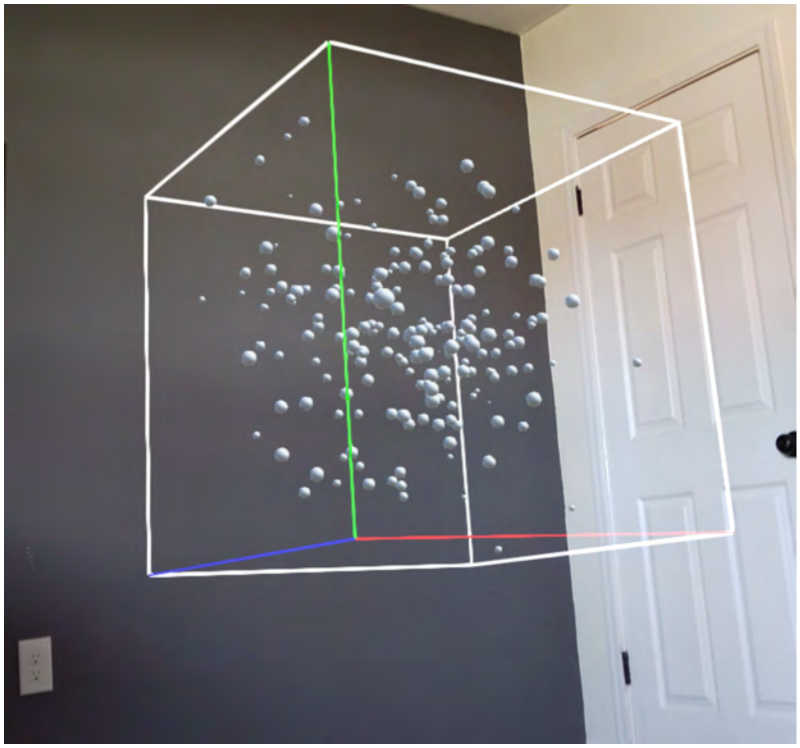

Figure 3.6: Scatterplot as Displayed on the HoloLens 2 AR Headset

example of network diagrams being used for 3D data visualizations [5]. Figure 3.7 and Figure 3.8 show an example how the network diagrams looked on the desktop computer and on the HoloLens 2 respectively.

None of the charts presented in this study included any kind of text labels, to keep the focus on the data encoded in the geometry of the charts rather than the context of that data.

Chart Data

All data sets participants viewed in the charts in this study were synthetically generated.

The data used in the network diagrams was generated using the Watts-Strogatz method to represent "small-world networks", because these networks closely model many real world data sets, including biological and social networks [94]. Networks

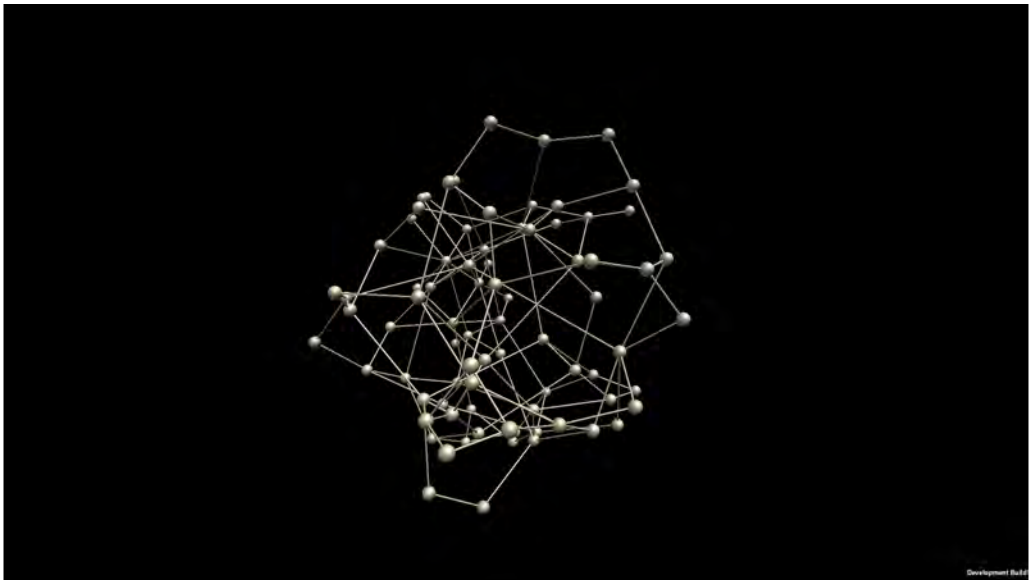

Figure 3.7: Network Diagram as Displayed on the Desktop Computer

Figure 3.8: Network Diagram as Displayed on the HoloLens 2 AR Headset

were all initialized with 80 nodes and 4 edges per node, and then rewired via the Watts-Strogatz algorithm with a swap probability of 72%.

The data for the bar charts and scatter plots was generated using the same technique as Whitlock et al. [96]. A data set consisting of Gaussian samples was generated, and then modified as described below.

For the scatter plot data, three Gaussian distributions were generated by summing repeated samples from a liner random distribution and then re-scaling using the Central Limit Theorum. Points were constructed by taking one sample from each of the three distributions to serve as the x, y, and z coordinates respectively, ensuring that the dimensions were independently distributed from each other. The fourth data attribute, which was to be encoded as point size in the charts, was then adjusted to either introduce an extrema or introduce a trend into the chart's data set. To introduce an extremum, the a Gaussian distribution of twice the needed size was generated. All values except one were then linearly randomly sampled from the subset of the Gaussian distribution below the threshold, while the one remaining value was linearly randomly sampled from above the threshold. A threshold of 0.82 was chosen, meaning all values in the modified data set except the extremum fell below the 82nd percentile of the Gaussian sample. To introduce a trend, a value for the 4th dimension of each data point was sampled from a Gaussian distribution and then multiplied by an "offset function" as described in [96].

Bar chart data sets were similarly created, although with a bar placed for each row and column on the chart and only the heights generated from a Gaussian distribution. Extrema and trends were each introduced into half of the bar chart data sets respectively using the same technique as described for the scatter plots.

3.2.3 Tasks

At a high level, we test two types of visual analysis tasks: open tasks and closed tasks. The specific questions asked for each of these tasks are shown in Table 3.1.

Open tasks do not have a single correct answer, and require the participants to evaluate the data they are given to determine a viable answer to a question. Guided Search was selected as the basis for the open tasks in this study, and has been previously used to encourage discourse and communication [76].

Closed tasks have a single correct answer, and require the participants to evaluate the data they are given to determine the correct answer to a question.

	Closed	Open
Scatter Plot	Which point is the largest? Is there a trend? If so, what direction is it going in?	Which of the three axes (directions representing a variable) do the points suggest is the most important? How many sub-groups (also known as clusters) would you divide these points into?
Bar Chart	Which bar is the tallest? Is there a trend? If so, what direction is it going in?	Find a repeating pattern in the heights of the bars. How many categories would you divide all the bars into based on height?
Network Diagram	What is the diameter of the graph (the longest connected chain of nodes you can find, without repeating any nodes)? Which node has the most connections?	Which node is the most important? How many sub-groups (also known as clusters) would you divide these nodes into?

Table 3.1: Task Questions

The open tasks were intentionally left ambiguous and up to interpretation, to encourage discussion among teammates.

3.3 Hardware

3.3.1 Desktop

For the desktop modality, a laptop was connected to a set of external peripherals, including a full mechanical keyboard, a three-button wired mouse, and a single 60Hz, 24 inch, 1080p, 16:9 monitor. This setup was chosen to mimic common scenarios in industry.

3.3.2 Augmented Reality

For the augmented reality modality, the Microsoft HoloLens 2 was chosen. The HoloLens is a state-of-the-art OST-HMD, featuring a 60Hz 2k resolution 3:2 display for each eye with approximately 2,500 points of light per degree in the user's 52 degree (measured diagonally) field of view [41]. This contrasts with several recent augmented reality studies, which chose HHDs to test with [72]. Perhaps the most obvious limitation of HHDs is that they must be held in the user's hand, thus taking away any opportunity for two-handed interactions with the virtual content [7]. Riedlinger et al. performed a recent study comparing an HHD to the HoloLens that produced mixed results for favoring each device, but used a first-generation HoloLens and an HHD (the Tango tables) that has more sensors than a commonplace mobile device would today [80]. Because of this potential interference with gesturing and communication, and because several other studies have already been done exploring HHDs, OST-HMDs were chosen for this study.

3.4 Software

The application used for the visualizations was a custom-developed Universal Windows Platform (UWP) app. The Universal Windows Platform was chosen because it is the primary format of app the HoloLens can run, and because this was a single application could be developed and then deployed to both the desktop and the HoloLens to save time, simplify development, and keep the experience as consistent as possible between the two modalities.

The application was developed in Unity 2019 and Visual Studio 2019 using the Microsoft Mixed Reality Toolkit (MRTK), and scripted completely in C#. The full code is publicly available on GitHub at https://github.com/IronicTugBoat22/masters-thesis.

To simplify development, both HoloLens were running identical but separate instances of the app. The app would be launched from the same, marked position on the floor for each participant study session, to minimize any offset between the visualizations participants were looking at. Correlation between the two devices' software logs would be made by using timestamps from the devices' synchronized internal clocks.

3.5 Participants

Participants were purposely chosen who knew each other to "*facilitate rich and smooth conversation*" [55]. This was additionally used as a precaution to lower the risk of COVID-19 spreading across study participants, as study participants were unable to socially distance from their teammates for the entire duration of the study. The only requirement placed on participants was that they be at least 18 years of age at the time of the study. Participants were recruited primarily through Arizona State University, and some through the community.


Figure 3.9: The Study Space

3.6 Study Space

The experiment was conducted indoors, in a room with a 15 x 15 ft space cleared of obstructions so that participants could walk around freely one one side, and some tables with the desktop computer set up on the other side. The lack of exterior windows was an intentional consideration to keep lighting consistent across trials. Figure 3.9 shows the cleared room used for the study space. The cords on the floor (taped over for safety) terminate in approximately the center of the space, and represents where the visualizations were centered in the augmented reality trials. The tape line to the bottom right marks the starting point participants used to start the visualization app, to ensure consistency of the visualization placement in the room between teams. Note that during the trails, the overhead lights were turned off, while the lights in the other, uncleared half of the room (to the left of the direction faced in Figure 3.9) were left on. This was not an intentional design choice, but rather left that way for consistency of lighting between trials.

3.7 Procedure

Participants were brought into the room by the study administrator and were walked through a verbal review of the study consent form. Participants were then given the opportunity to examine the consent form, before being required to sign it to continue with the study.

Participants then independently took the pre-study survey. Participants chose a number (either 1 or 2) for the purpose of later correlating their survey results, and entered this number as a part of the pre-study survey.

Participants would then run two sets of trials, one set with the desktop modality and one set with the augmented reality modality. All participants saw the same set of visualizations, but the order the participants ran the modalities in was randomized in advance between teams, to counter any learning effects. The order of the trials participants ran on each modality was similarly randomized for the same purpose. Participants would complete 24 total trials: 2 trials each for both of the 2 task types for each of the 3 chart types on both of the 2 modalities. Each trial used a unique visualization, so that a team never saw any given visualization more than once.

For the augmented reality modality, participants were each provided with a HoloLens headset. The study administrator instructed each participant in putting on their headset and verifying it was working. The study administrator then walked participants through calibrating the headset for their eyes using the on-board process on the HoloLens. Participants then each stood at the same marked position on the floor one at a time and launched the study app to ensure experimental conditions were as consistent as possible. A brief explanation on how users could change their view by physically walking and moving around the study space was given.

For the desktop modality, participants were both seated side-by-side in front of the one desktop monitor. A brief explanation of the mouse controls for changing the view was given.

The study administrator gave a brief description of the three charts participants would see in the trials: bar charts, scatter plots, and network diagrams. The study administrator then confirmed that both participants could see the correct visualization. Upon confirmation, the study administrator then read participants a question from the task questions in Table 3.1, instructed participants to "begin", and started a 2-minute timer.

Participants were allowed to freely navigate around the visualized chart (either via the mouse or walking around the room depending on the modality),and freely communicate with each other. Any questions participants asked of the study instructor were promptly answered. Questions on procedures were answered fully, while questions seeking clarification of intent for the task were kept as vague as possible to encourage participants to interpret and discuss the task execution.

After the two minutes were up, or the participants indicated they had decided on an answer, whichever happened first, the participants explained their answer to the study administrator, and justified it with some reasoning on how they reached their conclusion. Participants then returned to their starting positions (or reset the view on the desktop).

This process was repeated for each of the 24 trials, with participants completing 12 trials on their first modality before switching to the 12 trials on the second modality.

3.8 Recording

This study primarily relies on classification and and coding of user behavior, and as such falls under the "Qualitative Analysis" category proposed by Dünser et al.

Dünser et al. found this to be the second rarest category of AR user study evaluation methods [28]. Additionally, Lam et al. encourage evaluating interactions between participants, in addition to analyzing logs of collaborating users' interactions with the system, because the former doesn't generate traceable log entries [63].

To generate a rich set of data for exploratory analysis, three mechanisms were used to collect data during the study: software logging of participant interactions, qualitative encoding from a video recording of the trials, and surveying participants via a pre-study and post-study questionnaire.

3.8.1 Software Logging

User actions were directly recorded by the software. These actions were recorded as the location and orientation of the user's point-of-view (POV) in six degrees of freedom: three axes for position, and three axes for rotation. This captured data could then be used to extrapolate rotate, pan, and zoom actions users took when completing the user tasks. This design is inspired by the 3D visualization study performed by Büschel et al., although they used an external spatial tracking system to determine the position of the user's AR device [19]. The location and orientation were both sampled once per frame in the software, and thus with the software targeting 60 frames per second (fps) on each modality, the sampling rate was approximately 60 Hz.

3.8.2 Qualitative Video Encoding

User communication was also recorded, for the purpose of assessing and comparing how much teams communicated across the different modalities. Participants were videotaped, and utterances and gestures from each were recorded. Communication recording in this manner is common in augmented reality user studies [45] [59].

Five of the seven dependent variables in the study were qualitatively encoded from the video recording of each trial: number of gestures for each participant, number of utterances for each participant, the number of times one of the participants looked at the other, the number of times participants clearly made an effort to establish conversational grounding, and which participant controlled the mouse during the desktop trials.

Participant gestures were defined to include any hand or arm motions used to communicate with their teammate. As such, participant movements that existed solely for their own reference (such as pointing while counting to themselves) or that were not communication-related (such as adjusting the augmented reality headset or their masks) were not counted. Sequential gestures were counted by either the participant lowering their arms between gestures or by a change in the train of thought between gestures (as indicated by their verbal utterances). This allowed sequential repetition of the same gesture to only be counted as a single gesture.

Utterances were defined to include any verbal communication directed at communicating with the other study partner. As such, utterances that were primarily for the participant's own reference were not included, such as a participant counting to themselves. A participant counting while illustrating a point to their teammate would be counted however. "Filler words" (such as "um" and "uh") were also not counted. Consecutive utterances were counted by the expression of a complete thought. As such, a participant uttering a single sentence with a large pause in the middle while they thought, would still be considered a single utterance. Similarly, several consecutive sentences spoken together (such as describing the location of a mark on the chart) were counted as a single utterance, unless the teammate asked a question and the participant adjusted their train of thought to answer the question. Acknowledgement utterances, as used by Kraut et al. [59] were also counted as separate utterances, even if they were part of a single sentence as long as the remainder of the sentence was not directed at acknowledgement. No filtering was applied to identify only task-relevant communication; all communication was considered.

The times participants looked at each other were counted based on visible changes of the participant's head position to look at their teammate. Consecutive looks were only counted if the participant looked away from their teammate between; a single, continuous look would be counted as a single look regardless of the length of time it lasted. Cases where both teammates looked at each other at the same time would still be counted as two separate looks - one for each teammate. Head motions where a participant moved their head toward their teammate, but did not take their eyes off of the chart, were not counted. The intent of this counting was to track the number of times participants switched their focus between the task space and communication space, as Kiyokawa et al. [55] and Billinghurst et al. [9] noted this was significant for communication.

Conversational grounding was based on the work of [57] and counted based on concerted attempts between the teammates to ensure they had the same understanding of what each other were attempting to communicate. Two different instances of this were recorded: 1) Teammate A makes a statement. Teammate B asks a clarifying question about that statement. Teammate A answers the question. Teammate B acknowledges the answer. 2) Teammate A makes a gesture or statement and then asks Teammate B whether they understood. Teammate B then answers. Teammate A acknowledges the answer. These instances follow the Inter-referential Life Cycle model used by Chastine et al. [20].

Which participant controlled the mouse for each trial was simply recorded as one of three possible values: Teammate A, Teammate B, or both. "Both" was recorded whenever both teammates each controlled the mouse at some point during the trial.

The split between the length of time each participant controlled the mouse was not considered. Because the view port on the desktop trials was completely controlled by the mouse, recording the participant who controlled the mouse also recorded the participant who controlled the view port. This participant (or both in the case of "both") was labelled the "driver".

3.8.3 Participant Questionnaires

Bach and Scapin noted that user questionnaires are valuable in evaluating the functionality of, and cross-referencing with performance data for, mixed reality systems [2].

The pre-study questionnaire collected basic demographics about participants and asked participants to rate their familiarity with computers, augmented reality, and their study partner.

The post-study questionnaire was based on the NASA Task Load Index (TLX). The NASA TLX has been used in several studies, including [72], [92]. Despite this, Dey et al. found only one of the twelve augmented reality collaboration studies surveyed in their paper used the NASA TLX, as did only two of 71 augmented reality interaction studies they surveyed [25]. Because of this, six questions were included to line up with the six key areas of the NASA TLX: mental demand, physical demand, temporal demand, performance, effort, and frustration. An additional two free-response questions were included to provide participants an opportunity to give feedback on their experience with the devices used for the different modalities. One question asked which device participants preferred, and the other asked whether participants found communicating with their teammate easier on either device.

Chapter 4

RESULTS

Seven dependent variables were monitored during the study: task response time, number of gestures for each participant, number of utterances for each participant, the number of times one of the participants looked at the other, the number of times participants clearly made an effort to establish conversational grounding, which participant controlled the mouse during the desktop trials, and the distance participants moved the view port. These variables were monitored via a combination of visualization software logging and audio/video recording participants as described in Chapter 3. Correctness was not monitored for closed tasks because this was determined to be more relevant to a study focused on perception rather than collaboration, and was seen as secondary to rather than a contributor towards or indicator of collaboration styles.

All statistical analysis was performed with $p = 0.05$ to determine significance. Pairwise t-tests were generally employed, with single factor ANOVA being used as needed. Chi-squared independence tests were used sparingly as a sanity check to further validate results. The degrees of freedom, t-statistic, and p-value are reported for all tests.

4.1 Demographics

Several participant demographics were collected to lend credibility to the generalization of the study results, and to explore possibly relevant factors for context in the qualitative analysis of study results.

4.1.1 General Demographics

Three general demographics were collected: age, gender, and education level.

Participants ranged in age from 19 to 54 years old. Most participants (9) were in their 20's, while 1 was younger and the remaining 2 were older. The average age was 28 years old, and the standard deviation was 10.7 years.

Gender was fairly evenly represented, with 6 reporting male, 5 reporting female, and1 reporting nonbinary or genderqueer.

A diverse range of education levels was also represented, with 2 reporting "some college", 1 reporting "Associates degree", 4 reporting "Bachelors degree", 4 reporting a "Masters degree", and 1 reporting a "PhD" or "professional degree". No participants reported a not completing high school, a high school diploma, or trade school.

4.1.2 Background Factors

Four background factors were collected that the study designers felt may have an impact on the results: familiarity with computers, familiarity with augmented or virtual reality, familiarity with data visualization/data analysis, and teammate's familiarity with each other.

Participants were asked to rate how familiar they are with computers (desktop or laptop) on a scale of 1 to 5, with 1 representing "*Never touched one in my life*" and 5 representing "*I am a professional or hobbyist who works with computers frequently (IT/programming/building computers/etc)*". All participants reported at least a 3. Similarly, participants were asked to rate how familiar they are with augmented or virtual reality headsets, with 1 representing "*Never touched one in my life*" and 5 representing "*I use them frequently (professionally or for fun)*". The results for this question were nearly the reverse; Nearly all participants reported at most a 3, with

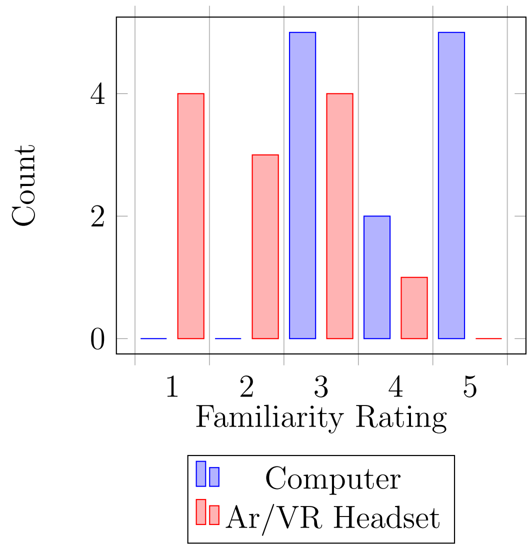

Figure 4.1: Participant Familiarity with Devices

only one participant reporting over a 3. The breakdown of the reported results are illustrated in Figure 4.1. For the desktop computer, 5 participants gave rated their familiarity a 3, 2 rated a 4, and 5 rated a 5. For the AR/VR headset, 4 rated a 1, 3 rated a 2, 4 rated a 3, and 1 rated a 4.

Participants also rated their familiarity with "*data visualization and/or data analysis*" on a scale from 1 to 5, with 1 representing "*I think I have heard those words before?*" and 5 representing "*I study these in university/college/trade school or I work professionally with these fields*". Responses here were distributed fairly evenly across the board, with each value receiving within a single participant difference from each other. These responses are illustrated in Figure 4.2.

Finally, participants were asked to rate how much they have worked with their study teammate before on a scale from 1 to 5, with 1 representing "*We just met today!*" and 5 representing "*We have known each other for many years, or have lived together for a while (best friends, childhood friends, partner/spouse, etc).*". All participants reported at least a 3, with 16.7% reporting a 3, and 41.7% each reporting a 4 and a 5. A skew towards participants being more familiar with each other was

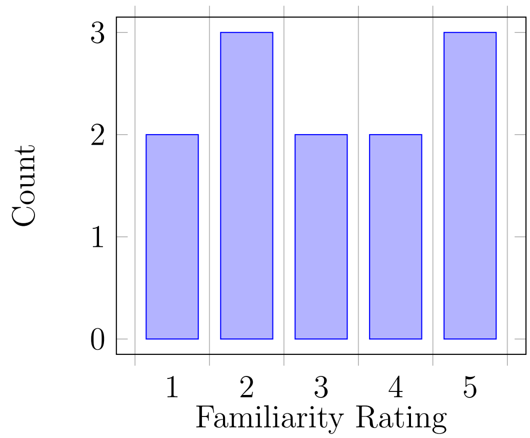

Figure 4.2: Participant Familiarity with Data/Info Visualization

expected due to participants selecting their own study teammates, and the study organizers purposely encouraging participants to select teammates they already had close contact with as a preventative measure to reduce the risk of COVID-19 spreading during the course of the study.

4.2 Task Completion

The task completion time, in seconds, was analyzed for influences from modality, chart type, and task type. Although not the focus of this study, the range of completion times (from a minimum of 2 seconds to the maximum allotted time of 120 seconds) gave confidence in a cross-sectional sample of task difficulties. It is worth noting as a possible limitation of this study that teams were cut off at two minutes, whether they had completed the task or not. As such, the actual task completion times may be significantly longer in some cases, but this would affect all conditions.

Modality and Task Completion Time

No significant change in task completion time was found when analyzed across modality using t-tests. A summary of the results can be seen in Table 4.1.

Constant Factors	Modality	Mean (s)	Std. Dev. (s)	t-stat	p-val
Bar Chart, Closed	Desktop HoloLens	62.4 58.3	29.1 31.0	t(11)=0.362	0.724
Bar Chart, Open	Desktop HoloLens	98.5 97.9	28.6 29.2	t(10)=0.0762	0.941
Network, Closed	Desktop HoloLens	112 105	18.8 22.4	t(9)=0.730	0.484
Network, Open	Desktop HoloLens	100. 87.7	21.1 32.2	t(10)=0.968	0.356
Scatterplot, Closed	Desktop HoloLens	88.0 80.0	35.7 37.8	t(10)=0.576	0.577
Scatterplot, Open	Desktop HoloLens	101 86.6	25.7 27.3	t(9)=1.27	0.235

Table 4.1: Pairwise T-Tests for Task Completion Time in Seconds Compared Across Modality

The mean task completion time was consistently shorter in the augmented reality modality, but the standard deviation (and thus variance) was also consistently higher, but neither difference was statistically significant. Therefor, Hypothesis 8: "*Participants will take significantly longer to complete tasks in the augmented reality modality than in the desktop display modality*" must be rejected. No statistically significant difference in task completion time was observed.

The difference in task completion time across modalities observed by Büschel et al. [19] is inconsistent with the results of this study. One possible explanation is that Büschel et al. worked with single participants rather than pairs of collaborating participants, and the effect they observed does not translate into multiple people synchronously collaborating on the same task. The same difference observed by Billinghurst et al. [9] is also inconsistent with this study. While Billinghurst et al. did work with pairs of collaborating participants, the fundamental nature of the tasks in their study is different. Their tasks were more focused on logical problem solving by manipulating playing pieces, while the tasks in this study were focused on analyzing data visualizations. The results from the two may not be directly comparable.

Chart Type and Task Completion Time

Running single factor ANOVA tests on the task time over different chart types, while holding the modality and task type constant, produced mixed results. A summary of these can be seen in Table 4.2.

Constant Factors	Chart Type	Mean (s)	Std. Dev. (s)	F-Stat	p-value
Desktop, Closed	BC	65.7	30.9	F(2,27)=6.28	0.00578
	ND	112	18.8		
	SP	84.8	35.9		
Desktop, Open	BC	96.4	29.3	F(2,27)=0.0801	0.923
	ND	99.0	21.9		
	SP	101	25.7		
HoloLens, Closed	BC	58.3	31.0	F(2,33)=7.80	0.00168
	ND	107	21.1		
	SP	76.5	38.0		
HoloLens, Open	BC	93.8	31.3	F(2,33)=0.133	0.876
	ND	90.4	32.1		
	SP	87.3	27.6		

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot

Table 4.2: Single Factor ANOVA for Task Completion Time in Seconds Compared Across Chart Type

Two of the four cases returned statistically significant results: the { Desktop Closed } condition (F(2,27)=6.28 & p=0.00578) and the { HoloLens Closed } condition (F(2,33)=7.80 & p=0.00168). Performing a one-tailed paired t-test in the { Desktop Closed } condition across the Bar Chart and Network Diagram (p=0.00425), across the Bar Chart and Scatterplot (p=0.0789), and across the Network Diagram and Scatterplot (p=0.0.0456) confirmed that the Network Diagram was the Chart Type with statistically different task completion times. The same tests on the { HoloLens Closed } condition reached a similar conclusion with p-values (p=0.000510), (p=0.0300), and (p=0.0203) respectively, suggesting that task completion times for all three Chart Types were statistically different.

These results suggest that teams took significantly longer to complete the closed tasks on the Network Diagram, regardless of whether or not participants were in the Desktop or AR modality. This is consistent with observations from the video recording, where participants would frequently struggle to complete one of the Network closed tasks: "*What is the diameter of the graph (the longest connected chain of nodes you can find, without repeating any nodes)?*". This task was unique in requiring that participants count a large number of objects while observing certain rules, and was frequently commented on by participants as being particularly difficult in the allotted time without being able to "mark off" already counted nodes. 7 of the 12 trials on this specific task (58.3%) took the full two minutes of allotted time, and nearly all of the runs with this task resulted in participants guessing at an answer while admitting they were only guessing. Because of these contextual details observed during the study, this finding is likely not significant due to any specific aspect of Network Diagrams, but rather due to the task being markedly more difficult compared to the other tasks used in the study. This is not considered a confounding variable to the study, however, as the study is primarily concerned with communication and collaboration styles and not task completion times.

Task Type and Task Completion Time

Similarly to Chart Type, running paired t-tests on the task time over the two task types, while holding the modality and chart type constant, produced mixed results. A summary of these can be seen in Table 4.3.

These results suggest that teams took significantly less time to complete the closed tasks on the Bar Chart. This is consistent with observations from the video recording, where participants would quickly arrive at the answer to the task: "*Which bar is the tallest?*". The difficulty of this task was significantly reduced due to a technical design

Constant Factors	Task Type	Mean (s)	Std. Dev. (s)	t-stat	p-value
Desktop, Bar Chart	Closed Open	63.5 98.5	30.2 28.6	t(10)= -2.74	0.0207
Desktop, Network	Closed Open	112 99.0	18.8 21.9	t(9)= 1.81	0.104
Desktop, Scatterplot	Closed Open	84.8 101	35.9 25.7	t(9)= -1.20	0.262
HoloLens, Bar Chart	Closed Open	58.3 93.8	31.0 31.3	t(11)= -2.85	0.0158
HoloLens, Network	Closed Open	107 90.4	21.1 32.1	t(11)= 1.99	0.0723
HoloLens, Scatterplot	Closed Open	76.5 87.3	38.0 27.6	t(11)= -0.689	0.505

Table 4.3: Pairwise T-Tests for Task Completion Time in Seconds Compared Across Task Type

decision in the code that produced the bar charts. Rather than keeping the edges of the bounding box of the bar chart at a constant scale relative to the size of the chart, the scale (and thus bars in the chart) were adjusted so that the tallest bar perfectly filled the bounding box. This resulted in participants quickly spotting this, and pointing out the tallest bar correctly, including in the shortest task time recorded in the study (2 seconds). Because of these contextual details observed during the study, this finding is likely not significant due to any specific aspect of Bar Charts specifically, but rather due to the task being markedly easier compared to the other tasks used in the study. Similarly to the chart type and task completion time analysis, this is not considered a confounding variable to the study as the study is primarily concerned with communication and collaboration styles and not task completion times.

4.3 Communication

After the study, the audio and video recording of each experiment was reviewed. Utterances and gestures from each participant were tagged and timestamped. These tags included the participant the utterance or gesture came from and a category of either "selection", "acknowledgement", or "clarification" based on the encoding used by Kraut et al. [59] and the Inter-referential Life Cycle model used by Chastine et al. [20]. The number of gestures in each category was then counted.

Communication recording in this manner is common in augmented reality user studies [45] [59].

4.3.1 Gestures

Modality and Gestures

Modality appeared to have a significant effect on how much participants used gestures to communicate. A Chi-Square Test of the effect of modality on the number of participant gestures returned a p-value of 0.041, suggesting a statistically significant influence between the two variables; the modality influenced the total number of gestures teams made while collaborating on problem-solving. To explore deeper, a series of paired t-tests were run to compare the differences in gesturing while varying the modality but holding the chart type and task type constant. A summary of these tests can be seen in Table 4.4.

A statistically significant difference was found in both bar chart cases ($p=0.0378$ and $p=0.0431$) and both scatterplot cases ($p=0.00002$ and $p=0.0204$). In all cases, the mean number of total gestures a team made per unit time increased when using the HoloLens compared to the desktop computer. The distributions can be seen in 4.3.

Because of this, Hypothesis 3: "*The amount of gestures participants use when communicating will change significantly between when using the augmented reality headsets and when using the desktop computer.*" fails to be rejected. While the

Constant Factors	Modality	Mean (G/m)	Std. Dev. (G/m)	t-stat	p-value
Bar Chart, Closed	Desktop HoloLens	4.45 14.0	3.18 14.7	t(11)= -2.36	0.0378
Bar Chart, Open	Desktop HoloLens	3.44 5.51	2.13 2.32	t(10)= -2.31	0.0431
Network, Closed	Desktop HoloLens	4.02 6.78	3.21 3.66	t(9)= -2.21	0.0541
Network, Open	Desktop HoloLens	4.58 7.58	2.88 2.62	t(10)= -2.21	0.0510
Scatterplot, Closed	Desktop HoloLens	3.76 9.44	2.67 2.39	t(10)= -7.34	0.00002
Scatterplot, Open	Desktop HoloLens	3.42 6.84	3.11 2.49	t(9)= -2.81	0.0204

G/m = Gestures per Minute

Table 4.4: Pairwise T-Tests for Number of Gestures per Minute Compared Across Modality

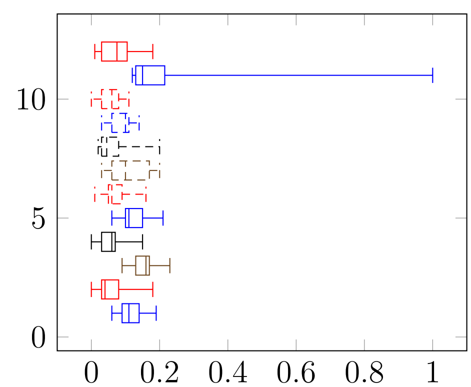

Figure 4.3: Number of Gestures per Second Across All Cases

results of the user study support a statistically significant influence between the two, the observed effect was in the opposite direction from the prediction.

One possible explanation for the apparent difference between the findings of Büschel et al. and observed influence in the opposite direction is a misunderstanding of the applicability of their results. In that study, participants expressed that physical objects provided context that helped orient themselves [19]. The key part there is “themselves”. The findings of this study do not contradict Büschel et al, but rather suggest that participants did not use the context of the physical objects to communicate orientation or position to their teammates even if they used that context to orient themselves. Rather than attempt to rely on descriptions of an object relative to the surrounding context, participants resorted to something considerably simpler and more effective when communicating with one another: simply pointing (gesturing) to what they were referring to.

It is worth noting that the number of gestures in the Desktop modality may be significantly under-counted due to the observational method employed by this study. A review of the video recording of the trials suggested that the participant controlling the mouse during the desktop trials would use the mouse to gesture to their teammate; particularly to point, or indicate "selection" as described by Kraut et al. [59]. These gestures were not counted, as the mouse pointer was not tracked and thus could not be correlated with the audio recording of the participants to determine gesturing. If these mouse gestures were also counted, the statistical significance of the influence between modality and number of participant gestures may not hold. Participants may gesture just as much between the two different modalities, although the mechanism of executing those gestures is clearly influenced.

In an attempt to explore this, the percentage of a team's total gestures made by each participant was calculated, and then the change in that percentage across modality was considered. This approach is similar to one used by Billinghurst et al. [9] and factors out individual differences where some participants naturally gesture more when communicating than other participants. Cases where neither participant gestured were considered to be a 50-50 split, since both participants contributed equally. A chi-squared test was then run on the modality and the percentage of the number of gestures each participant in the team gave. This test supports the idea that the distribution of the total gestures a team made between the teammates was influenced by the modality $\chi^2(24, N = 137) = 46.9, p = 0.016$. Further exploration with paired t-tests provided a less-clear picture, however, the results of which are summarized in Table 4.5.

Constant Factors	Modality	Mean	Std. Dev.	t-stat	p-value
Bar Chart, Closed	Desktop	26.0%	37.8%	t(11)= -2.22	0.0483
	HoloLens	51.1%	21.9%		
Bar Chart, Open	Desktop	37.3%	34.0%	t(10)= -1.15	0.275
	HoloLens	20.6%	20.6%		
Network, Closed	Desktop	49.3%	43.8%	t(9)= -0.316	0.759
	HoloLens	54.5%	28.1%		
Network, Open	Desktop	31.9%	35.3%	t(10)= -1.34	0.209
	HoloLens	51.1%	24.1%		
Scatterplot, Closed	Desktop	30.0%	32.1%	t(10)= -3.13	0.0107
	HoloLens	60.9%	18.4%		
Scatterplot, Open	Desktop	41.8%	37.8%	t(9)= -0.573	0.580
	HoloLens	49.8%	20.7%		

Table 4.5: Pairwise T-Tests for Percent of Gestures Made by a Single Teammate Compared Across Modality

Significant changes in the distribution of gestures between teammates were only observed in the {Bar Chart, Closed task} case (p=0.0483) and the {Scatterplot, Closed task} case (p=0.0107). This suggests that even if the mouse gestures were counted, the change in total gestures per unit time for a team across modalities would still be observed. If the un-counted mouse gestures would make a significant impact on that result, then the distribution of gestures between teammates should significantly change across modality, since only one of the two teammates operated the mouse in the overwhelming majority of desktop trials (only 10 of 65 desktop trials, 15.4%, had more than one participant take time operating the mouse). Because the distribution of counted gestures between teammates did not significantly change in most cases, there is likely another factor influencing the change in gestures aside from a failure to count the mouse-based gestures.

The results of this study provide evidence that the modality influences the distribution of the gestures between the two teammates on a team. Not only were the total number of gestures influenced, but the distribution of the total gestures made between the two teammates may have been influenced as well. The lack of counting gestures made with the mouse is consistent with this and a possible confound and plausible explanation that warrants further exploration, but the results also provide evidence that this was not the sole cause for the observed difference in gesture patterns between the desktop and augmented reality modalities.

Other Factors on Gestures

Neither task type nor chart type were found to influence either total gestures normalized for task completion time nor gesture distribution between teammates when paird t-tests and single factor ANOVA were performed respectively. The only significant change found was in the { HoloLens, Scatterplot } case when comparing across task types (p=0.0135), where participants gestured more per unit time with closed tasks (M=0.158, SD=0.0379) than with open tasks (M=0.107, SD=0.0414). No significant result was found in any of the other nine cases.

4.3.2 Utterances

The influence of modality on utterances was similarly examined with a chi-squared independence test. The p-values from these tests are summarized in Table 4.6. None of the independent variables (modality, chart type, or task type) were found to have a significant influence on any of the observed utterance measures. Further analysis of each case holding the other independent variables constant and then applying a paired t-test further supports this conclusion. A summary of the analysis can be seen for the impact of modality in Table 4.7, the impact of chart type in Table 4.8, and the impact of task type in Table 4.9. Similarly, summaries for the impact of modality, chart type, and task type on the distribution of gestures between participants can be seen in Table 4.10, Table 4.11, and Table 4.12 respectively. Hypothesis 4, "*Participants will verbally communicate less when using the augmented reality headsets than when using the desktop computer*" can thus be rejected. Put simply: the modality did not influence how much participants talked with one another.

	Distribution of Utterances Between Teammates	Total Utterances Normalized for Task Completion Time
Modality	0.394	0.776
Chart Type	0.370	0.422
Task Type	0.404	0.876

Table 4.6: P-values for Chi-Squared Tests on Factors Influencing Utterances

Constant Factors	Modality	Mean (U/m)	Std. Dev. (U/m)	t-stat	p-value
Bar Chart, Closed	Desktop HoloLens	18.2 25.2	6.09 20.8	t(11)= -1.05	0.316
Bar Chart, Open	Desktop HoloLens	14.2 14.5	2.59 6.79	t(10)= -0.160	0.876
Network, Closed	Desktop HoloLens	18.2 14.5	5.73 4.82	t(9)= 1.43	0.186
Network, Open	Desktop HoloLens	18.5 15.7	3.28 5.99	t(10)= 1.44	0.181
Scatterplot, Closed	Desktop HoloLens	18.7 18.4	4.48 4.37	t(10)= 0.408	0.692
Scatterplot, Open	Desktop HoloLens	17.3 19.7	5.21 4.71	t(9)= -0.930	0.377

U/m = Utterances per Minute

Table 4.7: Pairwise T-Tests for Number of Utterances per Minute Compared Across Modality

This finding at first appears to contradict the findings of Kraut et al. [58] and Gergle et al. [35], but some context provided by the qualitative analysis of the video recordings offers a plausible explanation for the apparent contradiction. Although the quantity of the participants' verbal communication was not influenced, the content was. Participants were observed frequently verbally communicating to ask their teammate (the one controlling the mouse) to change or manipulate the view of the

Constant Factors	Chart Type	Mean (U/m)	Std. Dev. (U/m)	F-Stat	p-value
Desktop, Closed	BC ND SP	18.6 18.2 18.7	6.35 5.73 4.72	F(2,27)=0.0215	0.979
Desktop, Open	BC ND SP	14.3 18.2 17.3	2.68 3.35 5.21	F(2,27)=2.75	0.0820
HoloLens, Closed	BC ND SP	25.2 14.0 18.8	20.8 4.99 4.41	F(2,27)=2.38	0.108
HoloLens, Open	BC ND SP	15.0 15.8 18.9	6.74 5.72 4.69	F(2,27)=1.50	0.239

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot, U/m = Utterances per Minute

Table 4.8: Single Factor ANOVA for Number of Utterances per Minute Compared Across Chart Type

Constant Factors	Task Type	Mean (U/m)	Std. Dev. (U/m)	t-stat	p-value
Desktop, Bar Chart	Closed Open	18.8 14.2	6.05 2.59	t(10)= 2.11	0.0609
Desktop, Network	Closed Open	18.2 18.2	5.73 3.35	t(9)= -0.0235	0.982
Desktop, Scatterplot	Closed Open	18.7 17.3	4.72 5.21	t(9)= 0.671	0.519
HoloLens, Bar Chart	Closed Open	25.2 15.0	20.8 6.74	t(11)= 1.56	0.147
HoloLens, Network	Closed Open	14.0 15.8	4.99 5.72	t(11)= -1.08	0.305
HoloLens, Scatterplot	Closed Open	18.8 18.9	4.41 4.69	t(11)= -0.0237	0.98

U/m = Utterances per Minute

Table 4.9: Pairwise T-Tests for Number of Utterances per Minute Compared Across Task Type

task space in the desktop modality. No communication about manipulating the view of the task space was observed in the augmented reality modality, as teammates each had their own headsets, and thus could independently control their own views of the

Constant Factors	Modality	Mean	Std. Dev.	t-stat	p-value
Bar Chart, Closed	Desktop	52.9%	8.82%	t(11)= 0.399	0.698
	HoloLens	52.0%	9.46%		
Bar Chart, Open	Desktop	53.2%	9.33%	t(10)= 1.60	0.140
	HoloLens	50.0%	10.6%		
Network, Closed	Desktop	57.8%	7.60%	t(9)= 0.256	0.803
	HoloLens	56.8%	11.1%		
Network, Open	Desktop	54.2%	8.40%	t(10)= -0.604	0.559
	HoloLens	56.0%	9.30%		
Scatterplot, Closed	Desktop	55.3%	10.1%	t(10)= 0.0452	0.965
	HoloLens	55.2%	8.01%		
Scatterplot, Open	Desktop	56.0%	13.3%	t(9)= 1.00	0.343
	HoloLens	53.5%	7.56%		

Table 4.10: Pairwise T-Tests for percent of Utterances made by a Single Teammate Compared Across Modality

Constant Factors	Chart Type	Mean	Std. Dev.	F-Stat	p-value
Desktop, Closed	BC	52.5%	9.40%	F(2,27)=0.956	0.397
	ND	57.8%	7.60%		
	SP	56.6%	9.63%		
Desktop, Open	BC	53.8%	9.66%	F(2,27)=0.108	0.898
	ND	55.1%	8.39%		
	SP	56.0%	13.3%		
HoloLens, Closed	BC	52.0%	9.46%	F(2,33)=0.375	0.690
	ND	53.8%	12.7%		
	SP	55.6%	7.76%		
HoloLens, Open	BC	49.3%	10.4%	F(2,33)=0.919	0.409
	ND	54.5%	10.3%		
	SP	51.9%	7.79%		

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot

Table 4.11: Single Factor ANOVA for Percent of Utterances Made by a Single Teammate Compared Across Chart Type

task space unlike the desktop modality where teammates shared a single view via the desktop monitor.

4.3.3 Space Switching

Modality appeared to have a significant effect on how frequently participants looked at their teammate. These looks were counted only when a participant phys-

Constant Factors	Task Type	Mean	Std. Dev.	t-stat	p-value
Desktop, Bar Chart	Closed Open	52.3% 53.2%	8.95% 9.33%	t(10)= -0.341	0.740
Desktop, Network	Closed Open	57.8% 55.1%	7.60% 8.39%	t(9)= 1.37	0.204
Desktop, Scatterplot	Closed Open	56.6% 56.0%	9.63% 13.3%	t(9)= 0.262	0.799
HoloLens, Bar Chart	Closed Open	52.0% 49.3%	9.46% 10.4%	t(11)= 0.990	0.343
HoloLens, Network	Closed Open	53.8% 54.5%	12.7% 10.2%	t(11)= -0.176	0.863
HoloLens, Scatterplot	Closed Open	55.6% 51.9%	7.76% 7.79%	t(11)= 1.70	0.118

Table 4.12: Pairwise T-Tests for Percent of Utterances Made by a Single Teammate Compared Across Task Type

ically and intentionally turned their head, and thus represent an intentional break from focusing on the task space to focusing on the communication space.

Modality and Space Switching

A significant increase in the frequency at which participants looked at their teammates in the augmented reality modality compared to the desktop modality was observed in 5 of 6 (83.3%) cases when holding Chart Type and Task Type constant. A summary of the paired t-tests run on these cases can be seen in Table 4.13.

Along a similar line of analysis as used with the task time, the one case { Network Closed } without a significant change in looks across modality (p=0.158) can be explained by the choice of task. One of the Network Closed tasks was: "*What is the diameter of the graph (the longest connected chain of nodes you can find, without repeating any nodes)?*". Analyzing the video recording of the trials using this task reveals that participants frequently used a "divide-and-conquer" approach to this task given its difficulty within the allotted 2 minute time limit. Participants were so absorbed in counting to themselves that they rarely looked at their teammate. This

Constant Factors	Modality	Mean (L/m)	Std. Dev. (L/m)	t-stat	p-value
Bar Chart, Closed	Desktop HoloLens	0.500 4.25	0.802 3.30	t(11)= -3.56	0.00445
Bar Chart, Open	Desktop HoloLens	0.873 3.49	1.09 3.16	t(10)= -2.51	0.0309
Network, Closed	Desktop HoloLens	0.960 1.80	2.21 1.06	t(9)= -1.54	0.158
Network, Open	Desktop HoloLens	1.25 5.40	1.66 3.09	t(10)= -4.31	0.00154
Scatterplot, Closed	Desktop HoloLens	0.436 3.82	0.852 2.70	t(10)= -3.69	0.00417
Scatterplot, Open	Desktop HoloLens	1.32 4.44	1.26 1.21	t(9)= -4.99	0.000750

L/m = Looks per Minute

Table 4.13: Pairwise T-Tests for Number of Looks per Minute Compared Across Modality

suggests that significantly mentally demanding tasks or tasks perceived as extremely difficult or impossible may reduce the influence modality has on how frequently participants look at their teammate, and suggests that future research can be conducted in this area to ascertain how task difficulty influences communication in team-analysis scenarios.

Chart Type and Space Switching

Running single factor ANOVA tests on the frequency participants looked at their teammate over different chart types, while holding the modality and task type constant did not find any significant difference. A summary of these can be seen in Table 4.14.

Task Type and Space Switching

Running paired t-tests on the frequency participants looked at their teammate over the two task types, while holding the modality and chart type constant produced

Constant Factors	Chart Type	Mean (L/m)	Std. Dev. (L/m)	F-Stat	p-value
Desktop, Closed	BC ND SP	0.480 0.960 0.300	0.839 2.21 0.762	F(2,27)=0.565	0.575
Desktop, Open	BC ND SP	0.960 1.32 1.32	1.10 1.74 1.26	F(2,27)=0.223	0.802
HoloLens, Closed	BC ND SP	4.25 1.90 3.65	3.30 1.02 2.64	F(2,33)=2.84	0.0728
HoloLens, Open	BC ND SP	3.50 5.15 4.30	3.01 3.07 1.37	F(2,33)=1.20	0.314

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot, L/m = Looks per Minute

Table 4.14: Single Factor ANOVA for Looks per Minute Compared Across Chart Type

mostly insignificant results. A summary of these can be seen in Table 4.15.

Constant Factors	Task Type	Mean (L/m)	Std. Dev. (L/m)	t-stat	p-value
Desktop, Bar Chart	Closed Open	0.545 0.873	0.825 1.09	t(10)= -0.855	0.412
Desktop, Network	Closed Open	0.960 1.32	2.21 1.74	t(9)= -0.732	0.483
Desktop, Scatterplot	Closed Open	0.300 1.32	0.762 1.26	t(9)= -2.68	0.0250
HoloLens, Bar Chart	Closed Open	4.25 3.50	3.30 3.01	t(11)= 0.558	0.588
HoloLens, Network	Closed Open	1.90 5.15	1.02 3.07	t(11)= -4.15	0.00162
HoloLens, Scatterplot	Closed Open	3.65 4.30	2.64 1.37	t(11)= -0.931	0.372

L/m = Looks per Minute

Table 4.15: Pairwise T-Tests for Looks per Minute Compared Across Task Type

Two of the cases produced a significant change in the frequency participants looked at their teammate. The { HoloLens Network } case (p=0.00162) can be explained by the same rationale as the analysis for the one inconsistent result in the look frequency over modality: the "divide-and-conquer" approach to the task "*What is the diameter of the graph (the longest connected chain of nodes you can find, without repeating any nodes)?*" resulted in participants being so absorbed in counting to themselves that they rarely looked at their teammate. This resulted in a significant difference between the open and closed task results for the Network chart type. It is unclear why this significant change does not hold true for the { Desktop Network } case here as well, as the difference between the frequency of participant looks across modality for { Network Open } was neither significant nor even in the correct direction to counter the result expected for the { Desktop Network } case here if the given explanation is correct.

It is also unclear why the { Desktop Scatterplot } case (p=0.0250) produced an inconsistent result with the majority of the cases here. These may be anomalies that would go away with a larger study size.

4.3.4 Conversational Grounding

Three forms of participants obviously attempting to establish conversational grounding were observed in the video recordings of the trials. The first, observed across both modalities, followed the pattern:

1. Participant A makes a statement or gesture.

2. Participant B asks a clarifying question about the statement or gesture.

3. Participant A provides an answer to the question.

4. Participant B acknowledges the answer to indicate understanding.

The second, also observed across both modalities, is a variation on the first:

1. Participant A makes a statement or gesture.
2. Participant A then asks if Participant B understood.
3. Participant B either confirms understanding or indicates confusion.
4. Participant A provides a clarification if confusion was indicated.
5. Participant B acknowledges the clarification to indicate understanding.

The third looked a little different, and was only observed in the augmented reality modality:

1. Participant A makes a statement, or explicitly calls for Participant B to come over.
2. Participant B moves so that their view is in alignment with Participant A's view.
3. Participant A makes or reiterates a statement.
4. Participant B acknowledges the statement to indicate understanding.

Reliably recognizing these sequences was difficult in the time the study was conducted in, largely due to limitations of the hardware used to collect the data, particularly the angle of the camera resulting in some difficult to see angles, the low-quality microphone resulting in some unclear dialog, and the requirement that participants wear face masks as a COVID-19 precaution obscuring mouths from the camera, making assessing which participant was speaking at times difficult. A more thorough review and analysis of the meaning behind the gestures and utterances should be conducted in future to provide a more reliable means to pinpoint moments of conversational grounding, and ascertain how the differing modalities impact the strategies participants use to establish conversational grounding.

Shifting Role of Communication

To pull apart and understand the findings of this study and how they line up with the findings from Kiyokawa et al. [55], a closer look at the role of conversational grounding in influencing communication must be taken. Kiyokawa et al. observed participants relying more on verbal communication when the one initiating the communication believed the participant being communicated to shared the same visual space. The one initiating the communication would more frequently use "deictic" phrases whenever they believed the other participant could see the position they were focused on. Similarly, the participant initiating the communication would use more "positional" phrases when they couldn't see a position. These observations can be described in terms of conversational grounding. Deictic phrases are phrases that cannot be understood from the speech alone, and includes words such as "this", "that", and "here" [55]. A deictic phrase must be put into context to have any meaning, such as the context of being accompanied by a gesture. This implies that the number of gestures would increase while the number of utterances would decrease as the participants are increasingly confident that they have a shared view and understanding of the task space - that they have reached conversational grounding. This implication, combined with the increased number of gestures participants used in the augmented reality modality when compared to the desktop observed by this study, provides some potential insight into the conversational grounding between teammates.

At first, the increased reliance on gestures in the augmented reality modality suggests that participants were more confident their teammate shared the same view of the task space when wearing the augmented reality headsets than when both looking at the same desktop monitor. The lack of a significant change in verbal communication (utterances) in the augmented reality modality observed in the study, however, suggests that the increased gestures observed were not taking the place of verbal communication; teammates were communicating more overall in augmented reality by augmenting their verbal communication with gestures. This does not contradict the findings of Kiyokawa et al. but does raise some questions about the reasoning behind how augmented reality impacts the necessity of verbal and nonverbal communication between collaborating teammates. Additional analysis of the content and purpose of both utterances and gestures should be performed. The role each plays in teammates' efforts to establish conversational grounding with each other may change between desktop and augmented reality.

4.4 Movement

Participant movement was captured both through video recording and through logging in the visualization software. For the purposes here, the software logging was used as the basis for analysis, while the video recording served as validation of the results.

4.4.1 Quantity of Movement

To determine the amount the participants moved a view port, the distance between each timestamped sample was calculated, and then all distances summed across the path for the trial. For comparison across modality, the average of the path distance for the two HoloLens's was compared to the path distance for the desktop.

A significant change in the average total distance participants moved their view port in the augmented reality modality compared to the desktop modality was only observed in 2 of 6 (33.3%) cases when holding Chart Type and Task Type constant. A summary of the paired t-tests run on these cases can be seen in Table 4.16. Note that the average distance the two participants in a team moved is considered for the augmented reality cases, to make it more comparable with the single viewport in the desktop modality.

Constant Factors	Modality	Mean (m)	Std. Dev. (m)	t-stat	p-value
Bar Chart, Closed	Desktop HoloLens	28.6 18.3	21.7 9.89	t(10)= 1.34	0.211
Bar Chart, Open	Desktop HoloLens	39.1 22.4	22.1 8.50	t(10)= 2.79	0.0191
Network, Closed	Desktop HoloLens	30.7 23.1	18.4 5.24	t(9)= 1.12	0.292
Network, Open	Desktop HoloLens	32.8 21.8	19.3 7.98	t(10)= 1.68	0.124
Scatterplot, Closed	Desktop HoloLens	33.2 19.5	15.3 9.10	t(10)= 2.41	0.0367
Scatterplot, Open	Desktop HoloLens	53.0 23.7	50.4 7.19	t(9)= 1.67	0.130

Table 4.16: Pairwise T-Tests for Distance View Port Traveled Compared Across Modality

These findings show mixed results for the effect modality has on the average distance participants move their view port. The means for all cases, including the ones with statistically significant differences, are consistently higher for the Desktop modality, but so is the standard deviation (and thus variance). With a lack of compelling evidence to support it, Hypothesis 2: "*The average number of participant movements will be higher when completing tasks with augmented reality than on a desktop computer*" must be rejected.

A notable limitation of this result, however, is in the fundamental difference between the two movement models used for the desktop and the HoloLens. While the desktop movement model, which uses rotate, zoom, and pan freely in 3D space, is commonly used in industry, the HoloLens movement observed from visualizing the view port paths of the participants in this study more closely aligns with the zoom, orbit, re-center model used by Büschel et al. [19]. The results of this study are inconclusive on the idea that HMDs will encourage more movement because they inherently afford participants more natural. It is not possible to make a confident assertion for nor rejection of this rationale, given these results and limitations. A similar study with a more consistent movement model, and a more refined definition for movement is worth pursuing.

No significant changes were found when comparing across Chart Type, and only one significant change was found when found when comparing across Task Type; the { Desktop Bar Char } case show participants moving more ($p=0.0225$) for open tasks ($M=39.1$, $SD=22.1$) than closed tasks ($M=28.6$, $SD=21.6$). Because this was the only significant case, there is not enough evidence to say any influence was found between Chart Type or Task Type and amount of movement.

4.4.2 Distribution of Movement

Similarly to gestures and utterances, the distribution of movements between teammates was calculated by dividing the amount a single teammate moved by the total movement across both teammates for the augmented reality modality cases.

No significant difference in the distribution of movements between participants was found when comparing across Chart Type and holding Task Type constant. A summary of the single factor ANOVA tests run on these cases can be seen in Table 4.17. Similarly, no difference was found when comparing across Task Type and holding Chart Type constant. A summary of the paired t-tests run on these cases can be seen in Table 4.18.

Constant Factors	Chart Type	Mean	Std. Dev.	F-Stat	p-value
HoloLens, Closed	BC	46.9%	11.9%	F(2,33)=0.602	0.553
	ND	50.0%	5.36%		
	SP	50.3%	6.73%		
HoloLens, Open	BC	50.3%	5.90%	F(2,33)=1.39	0.262
	ND	46.8%	4.74%		
	SP	51.4%	9.64%		

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot

Table 4.17: Single Factor ANOVA for Percent of View Port Distance Travelled by a Single Teammate Compared Across Chart Type

Constant Factors	Task Type	Mean	Std. Dev.	t-stat	p-value
HoloLens, Bar Chart	Closed	46.9%	11.9%	t(11)= -0.849	0.414
	Open	50.3%	5.90%		
HoloLens, Network	Closed	50.0%	5.63%	t(11)= 1.43	0.182
	Open	46.8%	4.74%		
HoloLens, Scatterplot	Closed	50.3%	6.73%	t(11)= -0.308	0.764
	Open	51.4%	9.64%		

Table 4.18: Pairwise T-Tests for Percent of View Port Distance Travelled by a Single Teammate Compared Across Task Type

Because no significant change was found between the distribution of movement made by each participant when comparing across Task Type in any cases, Hypothesis 1: "*The distribution of quantity of participant movements will not change significantly between tasks*" fails to be rejected. The observed participant behavior in this study is consistent with the findings of Büschel et al. [19].

4.4.3 Relative Position

The position coding used by Tang et al. [88] served as the inspiration for the position coding used in this study. They measured the position of the two participants each in one of six spaces around the visualization table as illustrated in Figure 4.4, and encoded the collective position of the two. This study presented in this thesis combined the idea of examining participant positions relative to each other and the significance of the overlap of communication and task spaces noted by Billinghurst

Figure 4.4: Positional Arrangements Encoded by Tang et al. [88]

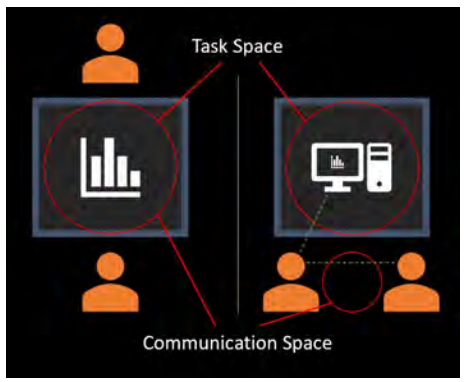

Figure 4.5: Communication and Task Spaces in Two Common Collaborative Workspace Setups

et al. [9]. Figure 4.5 shows two common collaborative layouts, one of which has an overlapping communication and task space, and the other of which does not.

To adapt the positional arrangement for the free-form 3D space in a way that measures overlap of the communication and task spaces, a triangle can be formed between the two participants and the chart. The task space exists centered around the chart, while the communication space exists between the two participants [9]. Thus, the angle formed by the line between participant A and participant B and the line between the same participant A and the chart can be used to examine the overlap of the two spaces for participant A (and vice-versa for participant B). See Figure 4.6 for an illustration of this.

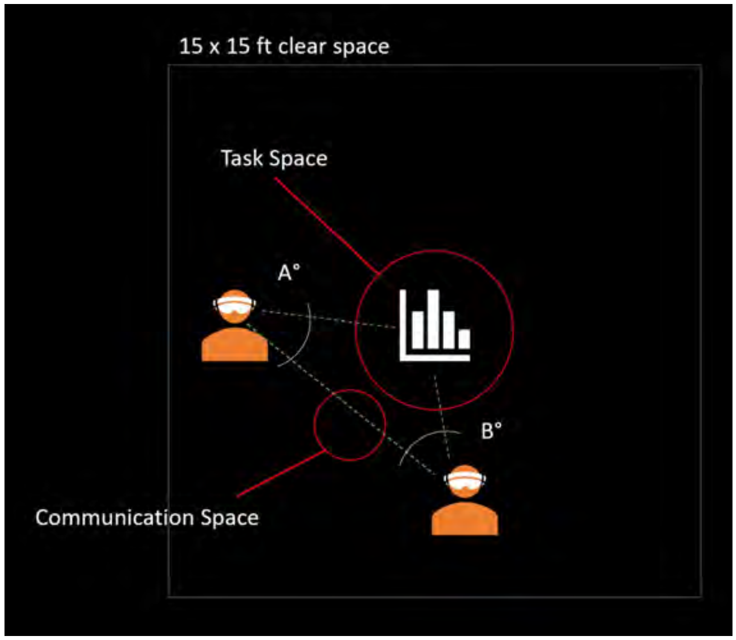

Figure 4.6: Positional Arrangements to Measure Overlap of Communication and Task Spaces

These angles were calculated at each sampled point in time from the positions of each participant as recorded by the software logging. 43.3 degrees was chosen as the threshold to separate "overlapping" from "non-overlapping" positional arrangements, from the perspective of each participant independently. This threshold was selected because it represents the horizontal viewing angle of the HoloLens (calculated from the reported 52 degree diagonal viewing angle on a 3:2 area), and thus is the greatest possible angle at which a participant could simultaneously see the chart and their teammate without turning their head. The total portion of time each participant spent below this threshold was then calculated, and then the two participants were averaged together. It is important to note that the two angles, one from each participant's point of view, are completely independent despite belonging to the same triangle, as no other restrictions on the third angle nor the length of any sides are placed. Thus one participant being below the threshold does not impact the other participant's ability to be below the threshold.

The mean amount of time participants spent below the threshold (meaning the communication and task spaces were overlapping from their point of view) across all 72 augmented reality modality trials was 37.0% with a standard deviation of 13.0% and a minimum of 13.2% and a maximum of 100.%. Because participants only spent a little over a third of their time on average with their communication and task spaces overlapping, Hypothesis 5: "*Participants will spend more time positioned such that their communication and task spaces overlap than not overlapping when using augmented reality*" must be rejected.

Based on observations of the video recordings, this could be a product of the cycles of communication people go through when collaborating, and role the relative positions of the communication and task spaces play in those cycles. Participants were observed, for example, to move through different behaviors of strategizing, independently gathering information, negotiating on different reasonings, and synchronizing their understandings. The task space and communication space may have different levels of importance in each of these different behaviors, thus motivating participants to focus more on one or the other rather than focusing on both all for the whole time. These behaviors are discussed in more detail in the Collaboration Styles section, but the role the two spaces have to play in the process of collaboration is worthy of additional research, as it may help inform the design of CSCW software in general, especially for remote and asynchronous collaboration and emerging modalities like augmented reality.

No significant difference was found in the amount of time participants spent with the communication and collaboration spaces overlapping when comparing across Chart Type nor Task Type while holding the other constant. A summary of the single factor ANOVA tests run on the Chart Type cases can be seen in Table 4.19. A summary of the paired t-tests run on the Task Type cases can be seen in Table 4.20.

Constant Factors	Chart Type	Mean	Std. Dev.	F-Stat	p-value
HoloLens, Closed	BC	41.7%	12.1%	F(2,33)=2.53	0.0953
	ND	31.9%	9.09%		
	SP	44.6%	19.9%		
HoloLens, Open	BC	30.3%	11.9%	F(2,33)=1.72	0.194
	ND	35.9%	10.5%		
	SP	7.7%	7.48%		

BC = Bar Chart, ND = Network Diagram, SP = Scatterplot

Table 4.19: Single Factor ANOVA for Average Percent of Time Participants Spent with Overlapping Task and Communication Spaces Compared Across Chart Type

Constant Factors	Task Type	Mean	Std. Dev.	t-stat	p-value
HoloLens Bar, Chart	Closed	41.7%	12.1%	t(11)= 2.18	0.0520
	Open	30.3%	11.9%		
HoloLens, Network	Closed	31.9%	9.09%	t(11)= -1.15	0.274
	Open	35.9%	10.5%		
HoloLens, Scatterplot	Closed	44.6%	19.9%	t(11)= 1.21	0.253
	Open	37.7%	7.48%		

Table 4.20: Pairwise T-Tests for Average Percent of Time Participants Spent With Overlapping Task and Communication Spaces Compared Across Task Type

4.5 Collaboration Styles

The collaboration styles described by Isenberg et al. [45] served as the basis for this study's own classification. The classification system used by Isenberg et al., however, was tested on a tabletop document-based system, and includes some styles that have definitions that do not translate into a collaborative augmented reality environment. To address this, an updated and generalized classification scheme is proposed based on the different participant behaviors observed through the recordings of the study sessions. These different behaviors are categorized based on the role they appear to play in the teammates' efforts on collaboration with each other.

- Strategizing - Participants would reason with each other about how to answer the question, and select a strategy for how they were going to determine the answer. This behavior was marked by one or both participants talking, and the content of their conversation focusing on interpreting the task at hand. Sometime participants would reference the visualization while in this behavior, and other times they would speak to each other independently of the visualization. This behavior was not exhibited in all trials, although when it was it would be either first or it would be second behind Independent Exploration.

- Independent Exploration - Participants would both be independently gathering information about the visualization. This behavior was marked by silence, and teammates moving and focusing with little or no regard for what the other was doing. This behavior was not exhibited in all trials, and seemed to be exhibited by some teams more than others. This was frequently either the first or second behavior seen during trials for the teams it was seen in. This behavior is similar to the "divide-and-conquer" behavior described by Tang et al. [88].

- Joint Exploration - Participants would gather information about the visualization together. This behavior was marked by on-going discussion, and teammates frequently calling each other over, or pointing out points on the visualization to each other.

- Reconciliation - Participants would take turns describing their understanding of the visualization and task, and attempt to reconcile that with their teammate's understanding. This behavior was marked by both participants talking, and included varying degrees of focus on each other and focus on the visualization. Participants would frequently attempt to establish conversational grounding. The degree of participation in this between teammates varied from team to team.

Further qualitative encoding will be necessary to fully address Hypotheses 6 and 7. Some discussion of each based on the observations from the video recordings follows.

Hypothesis 6 stated "*There will be no significant change in collaboration styles between augmented reality and desktop modalities*". The quantity of utterances stayed largely the same between the two modalities, but as noted in the Utterances section, the content of the discussion appeared to change between the modalities. Participants were more likely to communicate about the task at hand in the augmented reality modality, while a portion of their communication on the desktop modality was focused on directing the view port's movement. The quantity of gestures changed significantly between modalities, with participants gesturing more frequently in the augmented reality modality. From these measures, it is likely the collaboration style does change between the two modalities. For future work, a more formalized definition for these styles can be established, and the amount of time participants spend in each can be qualitatively coded similarly to the technique used by Isenberg et al. [45].

Hypothesis 7 stated "*There will be no significant change in collaboration styles between tasks*". The observations from the video recording produce mixed results towards this hypothesis. In general, both participants engaged in most trials, consistent with the observations of Tang et al. [88]. The only major exception to this was one of the questions in the { Network Closed } cases: "*What is the diameter of the graph (the longest connected chain of nodes you can find, without repeating any nodes)?*". Participants frequently resorted to the Independent Exploration behavior, and spent most of their time independently counting to themselves. As discussed in the Task Time and Modality and Space Switching sections, this task stood out as the one participants generally considered to be the most difficult. None were fully confident in the answer they gave, and frequently admitted to just guessing. This suggests that the perception of the difficulty of the task at hand may have more of an influence on the collaboration style than whether the task was open or closed. Future work into how the perception of the task and the experienced cognitive load using a measure such as the NASA Task Load Index (TLX) influences the collaboration style should be pursued.

4.6 Subjective Measures

The additional data collected directly from participant feedback through the exit survey at the conclusion of the study session provides some insight into how participants perceived working in each of the modalities.

4.6.1 NASA TLX

Participants directly rated their experience with the Desktop computer and the HoloLens augmented reality headset based on six key measures taken from the NASA Task Load Index (TLX) survey: mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants rated each modality on a scale from 1 to 7. A summary of the t-tests run on the differences between responses for each question compared across modality can be seen in Table 4.21.

TLX Measure	Modality	Mean	Std. Dev.	t-stat	p-value
Mental Demand	Desktop	4.92	1.88	$t(11) = 3.08$	0.0104
	HoloLens	3.58	1.17		
Physical Demand	Desktop	1.83	1.19	$t(11) = -5.06$	0.000365
	HoloLens	3.42	1.51		
Temporal Demand	Desktop	4.67	1.67	$t(11) = 1.82$	0.0969
	HoloLens	3.83	0.835		
Performance	Desktop	3.83	1.75	$t(11) = 0.261$	0.799
	HoloLens	3.67	1.87		
Effort	Desktop	5.50	1.57	$t(11) = 3.96$	0.00223
	HoloLens	3.58	1.38		
Frustration	Desktop	4.92	2.02	$t(11) = 4.70$	0.000651
	HoloLens	2.50	1.62		

Scores were on a scale of 1 to 7, with 1 = low, 4 = neutral, and 7 = high.

Table 4.21: Pairwise T-Tests for NASA TLX Survey Responses Compared Across Modality

Mental Demand

Participants responses showed a significant difference in their perception of mental demand (p=0.0104), rating the HoloLens (M=3.58, SD=1.17) as being less mentally demanding than the desktop (M=4.92, SD=1.88).

Physical Demand

Participants responses showed a significant difference in their perception of physical demand (p=0.000365), rating the HoloLens (M=3.42, SD=1.51) as more physically demanding than the desktop (M=1.83, SD=1.19). This result is understandable and expected; participants sat still while using the desktop, whereas they were required to physically walk around while using the HoloLens.

Temporal Demand

Temporal demand is how demanding a task was relative to the time requirement; it measures how rushed or hurried participants felt to accomplish the task on time. Participants responses didn't show a significant difference in their perception of temporal demand (p=0.0969) between the HoloLens (M=3.83, SD=0.835) and the desktop (M=4.67, SD=3.83). This is consistent with the analysis of task times, where no significant change was found in task completion time when comparing across modality, and suggests that participant's perception lined up with the measured result on this metric.

Performance

Participants responses didn't show a significant difference in their perception of performance (p=0.799) between the HoloLens (M=3.67, SD=1.87) and the desktop

(M=3.83, SD=1.75). Participants generally felt neutral leaning slightly positive towards their performance on each modality.

Effort

Participants responses showed a significant difference in their perception of effort (p=0.00223), rating the HoloLens (M=3.58, SD=1.38) as requiring less effort than the desktop (M=5.50, SD=1.57).

Frustration

Participants responses showed a significant difference in their perception of frustration (p=0.000651), rating the HoloLens (M=2.50, SD=1.62) as less frustrating than the desktop (M=4.92, SD=2.02). This is consistent with observations from the video recording of the trials, where participants would frequently comment on the difficulty of navigating around the visualization on the desktop.

4.6.2 Device Preference

The final two questions on the post-study survey asked participants about which device they preferred, and whether it was easier to communicate with their teammate while using one of the devices.

Participants overwhelmingly preferred the HoloLens to the desktop, with 11 of the 12 (91.7%) participants responding as such. 8 participants (66.7%) explicitly mentioned navigation or manipulating the view in their response. 3 participants mentioned the HoloLens being more "fun". The one participant who indicated preferring the desktop mentioned that coordination with their teammate was difficult on the HoloLens. It should be noted that there was a significant error in the alignment of the two participants' visualizations (estimated to be about 2 ft based on a review of the video recordings) for that participants' team, which may or may not have influenced their results. The other teammate still preferred the HoloLens despite this potential confound, however.

Participant feedback was decisively more mixed between devices when asked about which made communicating with their teammate easier. 7 participants (58.3%) said communication was eaiser while using the desktop. 4 participants (33.3%) said communication was easier while using the HoloLens. 1 participant (8.3%) said communication while using each device was just as easy. 5 participants, all of whom chose the desktop as easier to communicate with, explicitly mentioned ease of "pointing". These participants had a difficult time verifying they were referencing the same point in space as their teammate while using the HoloLens. This suggests that visible pointers, combined with increased accuracy of aligning up the visualizations across all teammates, is important to the success and ease of users' collaboration efforts.

4.7 Additional Observations

Chastine et al. observed users manipulating their own body or workspace to see the workspace from the same point of view as a teammate. The users also later explicitly stated that they were better able to communicate by observing the "exact view" of their teammate [20]. Participants were also observed reducing tasks in 3D space into 2D problems when using tablets to work with data visualizations in virtual reality [19]. The same behavior was observed in this study, with some teams explicitly stating they were trying to line the visualization up with a certain view to evaluate some property about the chart. This suggests that a mix of 2D and 3D visualizations of the same chart may be useful; the 3D visualization serves as the "overview" and provides a context from which participants can select a detailed view, to be displayed in 2D. This follows the overview-detail pattern discussed by Munzner [73]. Making this 2D detail view visible to all teammates would also satisfy some of the issues raised by participants and disucssed in the Device Preference section that frustrated establishing conversational grounding between teammates.

Chapter 5

CONCLUSION

Frost and Warren examined virtual reality's use in a collaborative architectural design process and concluded that virtual reality was useful and full of potential in collaboration. They also warned, however, that a proper motivation for usage and proper integration into the process is necessary, else virtual reality "*runs the risk of being just an expensive mean of presenting building projects*" [32]. With augmented reality devices often matching or exceeding the price of virtual reality, the search for proper motivation and integration into processes is just as, if not more important.

Based on the discussion in the Related Work section, an OST-HMD like the HoloLens would be best suited for a collaborative decision-making environment. As revealed through this study, however, some care must be taken when translating tasks between a traditional desktop computer and an augmented reality headset. While the total amount of communication appears to be unchanged, the content and role it plays between collaborating users appears to shift dramatically between the two modalities.

Different types of charts and different types of tasks did not seem to have a significant impact on participants' collaboration behaviors when using the augmented reality headsets.

5.1 Future Work

Several areas for future exploration have been identified through the course of this study.

5.1.1 Analyzing Communication Frameworks

A more thorough review and analysis of the meaning behind the gestures and utterances should be conducted in future to provide a more reliable means to pinpoint moments of conversational grounding, and ascertain how the differing modalities impact the strategies participants use to establish conversational grounding. From this, something like the Inter-referential Life Cycle model used by Chastine et al. [20] could be applied.

5.1.2 Increased Precision Positioning

One of the most unique aspects of augmented reality is that it allows users to more or less freely move about the physical space they are in while collaborating. It does not necessitate they stay relatively stationary like a traditional desktop computer does, nor does it fully obstruct the user's view like virtual reality does. Because of this, the way users position themselves in space relative to each other and the work space, and where participants are looking and focusing their attention can provide great insights into how they collaborate in augmented reality in ways not as easily afforded as in other modalities. More work should be done on precisely mapping teammate positions over time, as well as the direction teammates are looking. This can then be analyzed in the context of communication space and task space, and provide insight into the role each plays at different points during a team collaboration. A framework like the one applied by Isenberg et al. [45] could then be used to analyze collaboration styles.

5.1.3 Exploring Navigation Paradigms

The conclusion reached by Kiyokawa et al. that 3D tasks with spatial manipulations would be a place for further investigation with augmented reality interfaces [55] is echoed by this study. While an interaction and navigation paradigm (rotate, pan, zoom) described by Munzner for 2D visualizations [73], and common in industry for 3D applications was used for the desktop modality in this study, most participants provided feedback in the survey expressing that navigating with the HoloLens was much easier. The analysis of the participant movements paths during the augmented reality trials shows their movement patterns more naturally aliened with the zoom, orbit, re-center paradigm used by Büschel et al. [19]. A more detailed analysis of how and why participants move and interact with the visualization in the less-mediated modality of augmented reality may provide insight into ways to improve 3D navigation in visualizations displayed on traditional computer monitors.

5.1.4 Exploring Mixed Reality

Throughout the literature review and background research for this study, no examples of prior studies were found that used a viewing condition where participants are viewing and interacting with a mixed reality chart embedded in the real world. The "mixed reality" chart means that some aspect(s) of the chart are physically present in the real world; the chart is not purely virtual. Physical markers in the user's environment can benefit collaborating users by providing common references to facilitate coordinating actions [72]. These physical aspects of the chart are intended to serve as such common reference points that facilitate "conversational grounding" as described by Kraut et al [57], and follow the recommendation from Chastine et al. that multi-modal referencing techniques should be provided to augmented reality users

[20]. Additionally, this modality takes advantage of the opportunity for Decision-Visualization Environments to use multisensory experiences as noted by John et al. [51]. Further exploration into ways to bring a more true mixed reality experience to collaborating users should be investigated, as it could address some of the conversational grounding issues described in the participants' feedback on the post-study survey from this study.

5.2 Final Thoughts

Augmented reality has great potential and provides significant opportunities to revolutionize how users interact with computing devices and the data they contain. As seen in the studies from the past several decades reviewed as part of the research and inspiration for this study, augmented reality has had significant technical barriers to overcome before being a viable tool for data analysis and decision-making. The exploratory study presented here provides both some insight and cautionary notes on adapting information visualization from 2D monitors to augmented reality headsets. Taking the results and participant feedback of this study as just one more data point in a decades-long exploration of what augmented reality is and what can be done with it, the point has finally been reached where augmented reality is a viable, albeit expensive, collaboration tool. Decision visualization environments, such as the Decision Theater at ASU, have an exciting future ahead of them with the potential augmented reality brings to the area of collaborative decision-making!

REFERENCES

[1] Leila Alem and Jane Li. "A study of gestures in a video-mediated collaborative assembly task". In: *Advances in Human-Computer Interaction* 2011 (2011).

[2] Cédric Bach and Dominique L Scapin. "Obstacles and perspectives for evaluating mixed reality systems usability". In: *Acte du Workshop MIXER, IUI-CADUI*. Vol. 4. Citeseer. 2004.

[3] Kevin M Baird and Woodrow Barfield. "Evaluating the effectiveness of augmented reality displays for a manual assembly task". In: *Virtual Reality* 4.4 (1999), pp. 250–259.

[4] Patrick Baudisch and Ruth Rosenholtz. "Halo: a technique for visualizing off-screen objects". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2003, pp. 481–488.

[5] Daniel Belcher et al. "Using augmented reality for visualizing complex graphs in three dimensions". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE. 2003, pp. 84–93.

[6] Luıs MA Bettencourt et al. "Towards real time epidemiology: data assimilation, modeling and anomaly detection of health surveillance data streams". In: *NSF Workshop on Intelligence and Security Informatics*. Springer. 2007, pp. 79–90.

[7] Mark Billinghurst and Hirokazu Kato. "Collaborative mixed reality". In: *Proceedings of the First International Symposium on Mixed Reality*. 1999, pp. 261–284.

[8] Mark Billinghurst, Suzanne Weghorst, and Tom Furness. "Wearable computers for three dimensional CSCW". In: *Digest of Papers. First International Symposium on Wearable Computers*. IEEE. 1997, pp. 39–46.

[9] Mark Billinghurst et al. "Experiments with face-to-face collaborative AR interfaces". In: *Virtual Reality* 6.3 (2002), pp. 107–121.

[10] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.

[11] Frank A Biocca and Jannick P Rolland. "Virtual eyes can rearrange your body: Adaptation to visual displacement in see-through, head-mounted displays". In: *Presence* 7.3 (1998), pp. 262–277.

[12] Andrew C Boud et al. “Virtual reality and augmented reality as a training tool for assembly tasks”. In: *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*. IEEE. 1999, pp. 32–36.

[13] Tasseda Boukherroub, Sophie D’Amours, and Mikael Rönnqvist. “Toward decision theater design for community forest management & planning: the case of Québec”. In: *11th International Conference on Modeling, Optimization and Simulation (MOSIM’2016), Montréal, Québec, Canada*. 2016.

[14] Richard Brath. “3D InfoVis is here to stay: Deal with it”. In: *2014 IEEE VIS International Workshop on 3DVis (3DVis)*. IEEE. 2014, pp. 25–31.

[15] Brandon Bray. *What is Mixed Reality? - Mixed Reality*. URL: `https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality`.

[16] Matthew Brehmer and Tamara Munzner. “A multi-level typology of abstract visualization tasks”. In: *IEEE transactions on visualization and computer graphics* 19.12 (2013), pp. 2376–2385.

[17] Sabrina Bresciani and Martin J Eppler. “The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation”. In: *IEEE transactions on visualization and computer graphics* 15.6 (2009), pp. 1073–1080.

[18] Wolfgang Broll et al. “Arthur: A collaborative augmented environment for architectural design and urban planning”. In: *JVRB-Journal of Virtual Reality and Broadcasting* 1.1 (2004).

[19] Wolfgang Büschel et al. “Investigating the use of spatial interaction for 3d data visualization on mobile devices”. In: *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 2017, pp. 62–71.

[20] Jeffrey W Chastine et al. “Understanding the design space of referencing in collaborative augmented reality environments”. In: *Proceedings of graphics interface 2007*. 2007, pp. 207–214.

[21] Dimitris Chatzopoulos et al. “Mobile augmented reality survey: From where we are to where we go”. In: *Ieee Access* 5 (2017), pp. 6917–6950.

[22] Russell Chun. “Redundant Encoding in Data Visualizations: Assessing Perceptual Accuracy and Speed”. In: *Visual Communication Quarterly* 24.3 (2017), pp. 135–148.

[23] William S Cleveland and Robert McGill. “Graphical perception: The visual decoding of quantitative information on graphical displays of data”. In: *Journal of the Royal Statistical Society: Series A (General)* 150.3 (1987), pp. 192–210.

[24] William S Cleveland and Robert McGill. “Graphical perception: Theory, experimentation, and application to the development of graphical methods”. In: *Journal of the American statistical association* 79.387 (1984), pp. 531–554.

[25] Arindam Dey et al. “A systematic review of 10 years of augmented reality usability studies: 2005 to 2014”. In: *Frontiers in Robotics and AI* 5 (2018), p. 37.

[26] Kurt Chris Dohse. “Effects of field of view and stereo graphics on memory in immersive command and control”. In: (2007).

[27] Suyang Dong et al. “Collaborative visualization of engineering processes using tabletop augmented reality”. In: *Advances in Engineering Software* 55 (2013), pp. 45–55.

[28] Andreas Dünser, Raphaël Grasset, and Mark Billinghurst. *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand, 2008.

[29] Rob Edsall and Kelli L Larson. “Decision making in a virtual environment: Effectiveness of a semi-immersive “decision theater” in understanding and assessing human-environment interactions”. In: *Proceedings of AutoCarto*. Vol. 6. 2006, pp. 25–28.

[30] Clarence A Ellis, Simon J Gibbs, and Gail Rein. “Groupware: some issues and experiences”. In: *Communications of the ACM* 34.1 (1991), pp. 39–58.

[31] Niklas Elmqvist, Pierre Dragicevic, and Jean-Daniel Fekete. “Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation”. In: *IEEE transactions on Visualization and Computer Graphics* 14.6 (2008).

[32] Peter Frost and Peter Warren. “Virtual reality used in a collaborative architectural design process”. In: *2000 IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics*. IEEE. 2000, pp. 568–573.

[33] Susan R. Fussell, Leslie D. Setlock, and Elizabeth M. Parker. “Where Do Helpers Look? Gaze Targets during Collaborative Physical Tasks”. In: *CHI ’03 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003,

pp. 768–769. DOI: `10.1145/765891.765980`. URL: `https://doi.org/10.1145/765891.765980`.

[34] Joe L Gabbard and J Edward Swan II. “Usability engineering for augmented reality: Employing user-based studies to inform design”. In: *IEEE Transactions on visualization and computer graphics* 14.3 (2008), pp. 513–525.

[35] Darren Gergle, Robert E Kraut, and Susan R Fussell. “Action as language in a shared visual space”. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 2004, pp. 487–496.

[36] Irene Greif, ed. *Computer-Supported Cooperative Work: A Book of Readings*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[37] Irene Greif. *Computer-supported cooperative work: A book of readings*. Morgan Kaufmann Publishers Inc., 1988.

[38] Deirdre Hahn et al. “Does visualization affect perceptions of ethically complex policy decisions: an experimental study”. In: *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE. 2007, pp. 96–96.

[39] Jeffrey Heer and Michael Bostock. “Crowdsourcing graphical perception: using mechanical turk to assess visualization design”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010, pp. 203–212.

[40] Jeffrey Heer et al. “Creation and collaboration: Engaging new audiences for information visualization”. In: *Information visualization*. Springer, 2008, pp. 92–133.

[41] *HoloLens 2-Overview, Features, and Specs: Microsoft HoloLens*. URL: `https://www.microsoft.com/en-us/hololens/hardware`.

[42] Qian Hu et al. “Exploring the role of interactive computer simulations in public administration education”. In: *Journal of Public Affairs Education* 18.3 (2012), pp. 513–530.

[43] George P Huber. “Issues in the design of group decision support sytems”. In: *MIS quarterly* (1984), pp. 195–204.

[44] Kori Marie Inkpen. “Adapting the human-computer interface to support collaborative learning environments for children”. PhD thesis. University of British Columbia, 1997.

[45] Petra Isenberg et al. “Co-located collaborative visual analytics around a tabletop display”. In: *IEEE Transactions on visualization and Computer Graphics* 18.5 (2011), pp. 689–702.

[46] Petra Isenberg et al. “Collaborative visualization: Definition, challenges, and research agenda”. In: *Information Visualization* 10.4 (2011), pp. 310–326.

[47] Hiroshi Ishii, Minoru Kobayashi, and Kazuho Arita. “Iterative design of seamless collaboration media”. In: *Communications of the ACM* 37.8 (1994), pp. 83–97.

[48] Toru Ishikawa et al. “Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience”. In: *Journal of environmental psychology* 28.1 (2008), pp. 74–82.

[49] Yuta Itoh and Gudrun Klinker. “Light-field correction for spatial calibration of optical see-through head-mounted displays”. In: *IEEE Transactions on Visualization and Computer Graphics* 21.4 (2015), pp. 471–480.

[50] Robert Johansen et al. *Leading business teams: How teams can use technology and group process tools to enhance performance*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[51] Beatrice John et al. *Advancing decision-visualization environments: Empirically informed design guidelines*. 2018.

[52] Hassan A Karimi. *Telegeoinformatics: Location-based computing and services*. CRC press, 2004.

[53] David Kirk and Danae Stanton Fraser. “Comparing remote gesture technologies for supporting collaborative physical tasks”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, pp. 1191–1200.

[54] Naohiro Kishishita et al. “Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks”. In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2014, pp. 177–186.

[55] Kiyoshi Kiyokawa et al. “Communication behaviors of co-located users in collaborative AR interfaces”. In: *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE. 2002, pp. 139–148.

[56] Joshua M Knapp and Jack M Loomis. “Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 13.5 (2004), pp. 572–577.

[57] Robert E Kraut, Susan R Fussell, and Jane Siegel. "Visual information as a conversational resource in collaborative physical tasks". In: *Human–computer interaction* 18.1-2 (2003), pp. 13–49.

[58] Robert E Kraut, Darren Gergle, and Susan R Fussell. "The use of visual information in shared visual spaces: Informing the development of virtual co-presence". In: *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. 2002, pp. 31–40.

[59] Robert E Kraut, Mark D Miller, and Jane Siegel. "Collaboration in performance of physical tasks: Effects on outcomes and communication". In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*. 1996, pp. 57–66.

[60] Max Krichenbauer et al. "Augmented reality versus virtual reality for 3d object manipulation". In: *IEEE transactions on visualization and computer graphics* 24.2 (2017), pp. 1038–1048.

[61] Ernst Kruijff, J Edward Swan, and Steven Feiner. "Perceptual issues in augmented reality revisited". In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2010, pp. 3–12.

[62] Markos Kyritsis et al. "Issues and benefits of using 3D interfaces: visual and verbal tasks". In: *Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems*. 2013, pp. 241–245.

[63] Heidi Lam et al. "Empirical studies in information visualization: Seven scenarios". In: *IEEE transactions on visualization and computer graphics* 18.9 (2011), pp. 1520–1536.

[64] Jane Lessiter et al. "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory". In: *Presence: Teleoperators & Virtual Environments* 10.3 (2001), pp. 282–297.

[65] Yee Leung. *Intelligent spatial decision support systems*. Springer Science & Business Media, 2012.

[66] Shixia Liu et al. "A survey on information visualization: recent advances and challenges". In: *The Visual Computer* 30.12 (2014), pp. 1373–1393.

[67] Matthew Lombard and Theresa Ditton. "At the heart of it all: The concept of presence". In: *Journal of computer-mediated communication* 3.2 (1997), JCMC 321.

[68] K-L Ma. "Visualizing visualizations. User interfaces for managing and exploring scientific visualization data". In: *IEEE Computer Graphics and Applications* 20.5 (2000), pp. 16–19.

[69] Tim Marsh. "Staying there: an activity-based approach to narrative design and evaluation as an antidote to virtual corpsing". In: *Being there: Concepts, effects and measurement of user presence in synthetic environments, studies on new technologies and practices in communication series*. Citeseer. 2003.

[70] Mark McGill et al. "A dose of reality: Overcoming usability challenges in vr head-mounted displays". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, pp. 2143–2152.

[71] Jon Miller et al. *The Decision Theater: Collaborative Research Methodology*. 2019.

[72] Jens Müller et al. "A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices". In: *Proceedings of Mensch und Computer 2019*. 2019, pp. 399–410.

[73] Tamara Munzner. *Visualization analysis and design*. CRC press, 2014.

[74] Andrew YC Nee et al. "Augmented reality applications in design and manufacturing". In: *CIRP annals* 61.2 (2012), pp. 657–679.

[75] Jiazhi Ou et al. "Effects of Task Properties, Partner Actions, and Message Content on Eye Gaze Patterns in a Collaborative Task". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: Association for Computing Machinery, 2005, pp. 231–240. DOI: `10.1145/1054972.1055005`. URL: `https://doi.org/10.1145/1054972.1055005`.

[76] Kyoung S Park, Abhinav Kapoor, and Jason Leigh. "Lessons learned from employing multiple perspectives in a collaborative virtual environment for visualizing scientific data". In: *Proceedings of the third international conference on Collaborative virtual environments*. 2000, pp. 73–82.

[77] Long Qian et al. "Restoring the awareness in the occluded visual field for optical see-through head-mounted displays". In: *IEEE transactions on visualization and computer graphics* 24.11 (2018), pp. 2936–2946.

[78] Dheva Raja et al. "Exploring the benefits of immersion in abstract information visualization". In: *Proc. Immersive Projection Technology Workshop*. 2004, pp. 61–69.

[79] Patrick Renner and Thies Pfeiffer. “Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems”. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2017, pp. 186–194.

[80] Urs Riedlinger, Leif Oppermann, and Wolfgang Prinz. “Tango vs. HoloLens: A comparison of collaborative indoor AR visualisations using hand-held and hands-free devices”. In: *Multimodal Technologies and Interaction* 3.2 (2019), p. 23.

[81] Matthew W Rohrer. “Seeing is believing: the importance of visualization in manufacturing simulation”. In: *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)*. Vol. 2. IEEE. 2000, pp. 1211–1216.

[82] Dieter Schmalstieg et al. “The studierstube augmented reality project”. In: *Presence: Teleoperators & Virtual Environments* 11.1 (2002), pp. 33–54.

[83] Marc Aurel Schnabel et al. “From virtuality to reality and back”. In: (2007).

[84] Rebecca R Springmeyer, Meera M Blattner, and Nelson L Max. “A characterization of the scientific data analysis process”. In: *Proceedings of the 3rd conference on Visualization’92*. IEEE Computer Society Press. 1992, pp. 235–242.

[85] Mark St. John et al. “The use of 2D and 3D displays for shape-understanding versus relative-position tasks”. In: *Human Factors* 43.1 (2001), pp. 79–98.

[86] J Edward Swan and Joseph L Gabbard. “Survey of user-based experimentation in augmented reality”. In: *Proceedings of 1st International Conference on Virtual Reality*. Vol. 22. 2005, pp. 1–9.

[87] Zsolt Szalavári et al. ““Studierstube”: An environment for collaboration in augmented reality”. In: *Virtual Reality* 3.1 (1998), pp. 37–48.

[88] Anthony Tang et al. “Collaborative coupling over tabletop displays”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, pp. 1181–1190.

[89] Arthur Tang, Frank Biocca, and Lynette Lim. “Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study”. In: *Proceedings of PRESENCE* (2004), pp. 204–208.

[90] Arthur Tang et al. “Comparative effectiveness of augmented reality in object assembly”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2003, pp. 73–80.

[91] Melanie Tory et al. “Visualization task performance with 2D, 3D, and combination displays”. In: *IEEE transactions on visualization and computer graphics* 12.1 (2005), pp. 2–13.

[92] Xiangyu Wang and Phillip S Dunston. “Comparative effectiveness of mixed reality-based virtual environments in collaborative design”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.3 (2011), pp. 284–296.

[93] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[94] Duncan J Watts and Steven H Strogatz. “Collective dynamics of ‘small-world’ networks”. In: *nature* 393.6684 (1998), pp. 440–442.

[95] Dave D White et al. “Water management decision makers’ evaluations of uncertainty in a decision support system: the case of WaterSim in the Decision Theater”. In: *Journal of Environmental Planning and Management* 58.4 (2015), pp. 616–630.

[96] Matt Whitlock, Stephen Smart, and Danielle Albers Szafir. “Graphical Perception for Immersive Analytics”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2020, pp. 616–625.

[97] Daniel Wigdor et al. “Perception of elementary graphical elements in tabletop and multi-surface environments”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2007, pp. 473–482.

[98] Ji Soo Yi, Youn ah Kang, and John Stasko. “Toward a deeper understanding of the role of interaction in information visualization”. In: *IEEE transactions on visualization and computer graphics* 13.6 (2007), pp. 1224–1231.