Int. J. Human–Computer Studies 209 (2026) 103725

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

An epistemic network analysis of communication strategies during drawing-supported spatial dialogue in VR

Brandon Victor Syiem[a,b,*], Selen Türkay[b], Cael Gallagher[b], Christoph Schrank[c]

[a] School of Computer Science, The University of Sydney, Camperdown, Sydney, 2006, New South Wales, Australia
[b] School of Computer Science, Queensland University of Technology, Brisbane, 4000, Queensland, Australia
[c] School of Earth and Atmospheric Sciences, Queensland University of Technology, Brisbane, 4000, Queensland, Australia

HIGHLIGHTS

- We investigate the effects of 2D and 3D drawing modalities on supporting spatial dialogue in virtual reality collaborations.
- We use Epistemic Network Analysis (ENA) to understand the interconnections between user actions and speech using different drawing modalities during spatial dialogue in virtual reality.
- We find differences in communication strategies, and highlight the advantages and disadvantages of the chosen drawing modalities.
- We discuss potential solutions that leverage both 2D and 3D drawing capabilities in future VR applications to support spatial dialogue.

ARTICLE INFO

Keywords:
Virtual reality
Collaboration
Communication
Spatial dialogue
Gestures
2D drawing
3D drawing

ABSTRACT

Communicating spatial information is challenging using solely verbal or written language, and is often supported by non-verbal gestures and illustrative drawings. However, the growing need for communicating increasingly complex spatial information, coupled with the rise of remote collaboration, presents challenges that current screen-based solutions are ill-equipped to address. Virtual Reality (VR) offers capabilities to support both non-verbal gestures and complex visual aids, through embodied avatars and 3D virtual representations. However, the novelty of creating, referencing, and viewing 3D drawings in VR may influence the interlocutors' actions, speech and communication performance. We conducted a mixed-methods within-subject study with dyads to investigate the effects of drawing dimension (2D or 3D drawings) on spatial dialogue behaviours in VR. We found no significant effects of drawing dimension on communication performance and workload, but found significantly different interlocutor actions and speech. We discuss relevant implications and highlight considerations unique to the different communication strategies observed during 2D and 3D drawing use for supporting spatial dialogue in VR.

1. Introduction

In many professional fields such as architecture, navigation, geoscience, construction, and chemistry, effective collaboration often revolves around communicating spatial concepts. For instance, an architect discussing the design plans for a new building with a structural engineer, or a geologist describing the unique alignment of rocks observed during a recent field trip to their colleague, relies heavily on effectively articulating the placement of objects within space and their spatial relationships. Such forms of communication often occur in dialogic situations (Coventry et al., 2009; Tenbrink et al., 2017), and involve the interplay of *verbal descriptions, gestures, and visual aids like drawings and models* (Kang et al., 2015; Tversky et al., 2005). As such, spatial dialogue depends on a shared understanding of spatial reference frames and the ability to interact with physical representations, which enable interlocutors to ground their communication in a common spatial context (Coventry et al., 2009).

However, as remote collaboration becomes more common, particularly with the rise of distributed workforces and global projects, professionals increasingly find themselves engaging in these complex spatial discussions through online platforms that are limited to 2D screens. This shift to online environments places constraints on the

* Corresponding author at: School of Computer Science, The University of Sydney, Camperdown, Sydney, 2006, New South Wales, Australia.
Email address: brandon.syiem@sydney.edu.au (B.V. Syiem).

https://doi.org/10.1016/j.ijhcs.2025.103725
Received 20 January 2025; Received in revised form 26 August 2025; Accepted 31 December 2025
Available online 3 January 2026

natural multimodal interactions that are fundamental to spatial dialogue (Clark, 1991). Gestures, spatial organization, and the use of physical props or drawings are all difficult to replicate on a flat screen, forcing interlocutors to adopt language and actions that are atypical in face-to-face interactions. This mismatch between the affordances of face-to-face (F2F) spatial dialogue and the restrictions of 2D online meetings can lead to misunderstandings, inefficiencies, and increased cognitive load during collaboration (Wang et al., 2019).

Virtual reality (VR) offers potential solutions to the challenges associated with online spatial dialogue (Olaosebikan et al., 2022). VR can enable remote and co-located users to embody virtual avatars, and inhabit the same virtual environment. This enables the use of familiar gestures and spatial organizations around shared spatial reference frames during spatial dialogue (Osborne et al., 2023). VR also affords 3D capabilities, enabling systems that can augment spatial dialogue with visual illustrations that extend the capabilities of current F2F and 2D solutions. A prime example of this is the ability to create and update 3D illustrations during real-time communication, which is not easily achievable using physical props or 2D interfaces. Arguably the most universal form of real-time illustrations used during spatial dialogue is drawings. VR extends the capabilities of illustrative drawing by enabling 3D drawings. Unlike 2D drawings, which are limited to a 2D surface or plane, 3D drawings enable interlocutors to leverage the entire 3D virtual space for communication. However, the use of 3D drawings for communication may require users to adopt new actions/gestures and speech to fully leverage its capabilities, such as changing location while communicating to reference different perspectives of the 3D drawing. Additionally, familiarity with 2D drawings may offer distinct advantages over the expressiveness of 3D drawings, thereby influencing communication performance and perceptions differently. Given such differences, our study investigates whether the familiarity of supporting communication with 2D drawings constrains the use of novel 3D drawings during spatial dialogue in VR, leading to similar communicative strategies, perceptions and performance across both modalities. Specifically, we aim to answer the following research questions:

- ***RQ1:*** *How are interlocutors' actions and spoken language shaped by the use of 2D and 3D VR drawings during spatial dialogue about abstract 3D concepts?*
- ***RQ2:*** *How are communication performance and perceived workload influenced by the use of 2D and 3D drawings in VR during spatial dialogue about abstract 3D concepts?*

To address these questions, we conducted a within-subject user study with 10 dyads using a custom-built VR application. Our study investigates how 2D and 3D supporting VR drawings mediate interlocutor communicative actions and speech during spatial dialogue tasks related to structural geology. Structural geology was chosen due to the spatial complexity of the concepts involved, the need to relate abstract 3D concepts to environmental contexts, and the research team's expertise in the subject matter. We collected 7 h and 23 min of video and audio data of participants engaged in dialogue in VR, and conducted brief semi-structured interviews for supplementary data on user experience, preference for 2D or 3D drawing tools, perceived differences in communication strategies used between 2D and 3D drawing modalities, perceived workload, and task performance. While *not the primary focus of our study*, we also administered post-experiment tests and surveys to gain preliminary insights into the differences in understanding and perceived communication effectiveness between the two modalities. We employed open coding to generate codes related to user *actions* that mediate speech from the audio and video data. We then used epistemic network analysis (ENA) (Shaffer et al., 2016) to identify connections between coded elements and investigate differences in the underlying structure and strength of these connections between our conditions (communicating using 2D and 3D drawing).

We found that interlocutors employed significantly different communication strategies when using 2D or 3D drawings to support spatial dialogue in VR, with each offering unique perspectives and challenges. In the 2D condition, interlocutors frequently used additional drawings and gestures to *show* and *relate* multiple orthographic views. These interconnected actions in our findings highlight the need for solutions that not only enable multiple persistent drawings through larger/infinite 2D canvases, as seen in prior work (Grønbæk et al., 2024), but also provide effective means of locating and referring to created drawings on a large 2D canvas. In contrast, 3D drawings fostered a higher degree of spatial awareness, with our findings highlighting more frequent connections between actions that situate, and refer to, drawings within the relevant 3D environment. However, stronger connections between actions to create drawings and spatial reorganization of interlocutors spotlight challenges related to aligning different 3D perspectives between interlocutors. Additionally, heightened awareness of the surrounding space when using 3D drawing tools may necessitate careful design of the virtual environment to prevent dialogue irrelevant distractions, diminishing the spontaneous capabilities of drawing to support communication. Despite these differences between the drawing modalities, no significant effects were found in task completion times, time spent on drawings, perceived workload, or content understanding. Participants, however, expressed mixed preferences during our interviews. 2D drawings were quoted as allowing familiar strategies of creating, viewing and spatially organizing around the 2D drawing canvas. In contrast, participants highlighted the flexibility of 3D drawings, enabling them to create complex 3D drawings, meaningfully viewable from different spatial perspectives, while maintaining the ability to draw simpler 2D illustrations when needed.

Our findings highlight the unique communication strategies and interconnection of actions that interlocutors employ when supporting spatial dialogue with 2D or 3D drawings in VR. We find that while 2D drawings offer a familiar and structured approach to spatial dialogue, 3D drawings afford a richer, more immersive spatial experience. We discuss the implications of our findings, highlighting possibilities for supporting spatial dialogue by blending the affordances of 2D and 3D illustrations. We further discuss the importance of considering the potential cascading influences of enabling new drawing interactions on the interconnected actions that interlocutors employ during spatial dialogue in VR. Our findings contribute new insights on the influence of drawing dimensionality on spatial dialogue in VR and inform the design of future solutions for supporting spatial dialogue.

2. Related work

2.1. Spatial dialogue, gestures & drawings

Communicating spatial concepts often relies on shared spatial reference frames that serve as anchors for spatial language (Coventry et al., 2009). As such language evolved well before the advent of online meetings. Spatial reference frames and referent objects in face-to-face interactions are typically grounded in the physical context of the interlocutors. This shared physical context allows for the use of referential or deictic gestures (e.g., pointing) to facilitate spatial dialogue. Furthermore, evidence suggests that the use of such gestures predates vocal communication (Pollick and De Waal, 2007), highlighting how our linguistic systems have evolved to rely on visual representations (objects and environment) and visual cues (gestures) (Hayward and Tarr, 1995).

This coupling of visual and vocal modalities for spatial dialogue is also observed in discussions focusing on spatial and abstract concepts where referential actions, such as pointing, may fail. In such scenarios, humans have evolved to use iconic gestures to visually represent remote or abstract referents (Prieur et al., 2020). Modern gestures encompass a range of illustrations to complement spoken language, including the depiction of movement and the emphasis of pertinent words during speech. The prevalence of such diverse gestures to illustrate concepts in communication has led to a rich body of research detailing their differences and characteristics (McNeill, 1992; Ekman and Friesen, 1969; Efron, 1941). Despite the importance of gestures in communication, they are rarely used in isolation, and typically accompany other modalities of communication such as spoken words (Coventry et al., 2009) or facial expressions (Pollick and De Waal, 2007).

In contemporary work practices, the need to convey increasingly complex spatial ideas, structures, and relationships has grown. As a result, verbal language and gestures (whether iconic or deictic) are often insufficient. We now rely on a variety of physical and digital models and images (or props), and/or drawings to ground conversations around common frames of reference (Tversky et al., 2005). This reliance is evinced by the plethora of visual communication tools, such as whiteboards, projectors, screens, scale models, pin-boards, etc. commonly found in modern workspaces designed for collaboration. Such tools are particularly salient in communications related to spatial concepts, where interlocutors often prepare materials like blueprints, digital 3D models, or scale models to facilitate effective communication.

Drawings, in particular, serve as powerful tools for describing space and spatial relationships. Unlike other complementary modalities, such as 3D models and props, drawings allow for spontaneity (Oti and Crilly, 2021) and real-time refinement of mental models (Suwa and Tversky, 1996) (design formulation), which eludes strict requirements of prior preparation (like building/rendering a 3D model). Extensive prior work has detailed the use and cognitive processes behind the creation of artefacts using freehand 2D drawings in non-collaborative settings (Suwa and Tversky, 1996; Tversky, 2002; Pache, 2005; Pache et al., 2001). Additionally, relevant theoretical framings, such as image-enabled discourse (Snyder, 2011), emphasize the importance of the *spontaneous activity* of drawing in lending meaning to the visual artefact created through drawing for communication (Snyder, 2013, 2014). However, drawings have traditionally been limited to 2D representations, despite frequently attempting to depict 3D constructs (Arora et al., 2023; Oti and Crilly, 2021). This limitation forces interlocutors to project 3D concepts onto 2D surfaces, restricting the use of natural movements and spatial perspectives, and requiring multiple 2D drawings to illustrate different viewpoints. While the use of such projections has become standard across various fields, particularly those concerned with 3D structures and concepts, the formalization of these techniques is relatively recent in the history of drawings, originating in the 17th century to better communicate technical machines and instruments (Rovida, 2012).

VR technology provides capabilities to extend current 2D visual illustrations by enabling direct representations of 3D structures in space (Kingsley et al., 2019). Extensive prior work has explored methods for enabling the *creation* of 3D models and drawings in VR (Oti and Crilly, 2021; Jang et al., 2017; Israel et al., 2009). For instance, Jang et al. (2017) developed novel interaction methods for 3D drawing in VR that extend users' physical reach. Beyond the *creation* of 3D drawings, prior work has also demonstrated and discussed the potential of 3D illustrations in *supporting* visual thinking, communication, and conceptual design in fields such as architecture (Oti and Crilly, 2021), product design (Israel et al., 2009), and engineering (Pache, 2005). However, limited research has investigated the use of VR drawings for *communicating spatial concepts* in real-time multi-user setups, with prior VR drawing-based communication studies employing single-user VR systems (Oti and Crilly, 2021). Additionally, while 3D VR drawings offer potential advantages in communicating spatial concepts, they present a paradigm shift from the familiar and learned methods of communicating using 2D drawings. These differences between 2D and 3D drawing affordances in VR may lead interlocutors to adopt distinct communication strategies, which could inform the design of future technologies aimed at improving spatial dialogue in VR.

2.2. Multimodal communication in collaborative VR

The rise in remote collaborative work practices in recent years (Gifford, 2022) has intensified research on novel technological alternatives to screen-based video and teleconferencing systems for collaboration (Schäfer et al., 2022). Numerous studies have examined the use of virtual, augmented and mixed reality (VR/AR/MR) for collaboration (Schäfer et al., 2022; de Belen et al., 2019; Ens et al., 2019; Ouverson and Gilbert, 2021). Researchers have shown particular interest in how users communicate when collaborating remotely using VR (Li et al., 2021; Wei et al., 2022; Dwivedi et al., 2024). This interest stems from the promising capabilities of VR in enhancing distributed collaboration through embodied actions and spatialized experiences (Olaosebikan et al., 2022; McVeigh-Schultz et al., 2019; Piumsomboon et al., 2017), enabling forms of communication that combine non-verbal gestures, spatial positioning, and speech, which are unavailable in screen-based solutions.

Prior works have highlighted the use of deictic and body gestures to aid communication in collaborative systems that support embodied representations of the interlocutors (Olaosebikan et al., 2022; Gasques et al., 2021; Smith and Neff, 2018). Embodied representations also prompt users to be aware of their spatial position relative to others and task objects. For instance, Irlitti et al. (2024) found that users would reposition themselves to respect a remote collaborator's personal space based on their collaborator's avatar position in the MR space. Moreover, prior work has demonstrated how dyads were aware of, and considered different spatial perspectives during spatial dialogue in VR owing to the sense of 'spatial presence' (perception of existing within a space) that can be evoked in VR simulations (Uz-Bilgin et al., 2020). Most notably, the work presented by Smith and Neff (2018) highlights 'remarkable' similarities in the use of various gestures and spoken language during spatial dialogue related tasks between face-to-face and embodied VR conditions. These properties of communicating in VR suggest differences in the ways interlocutors act and speak when compared to communications enabled through screen-based mediums.

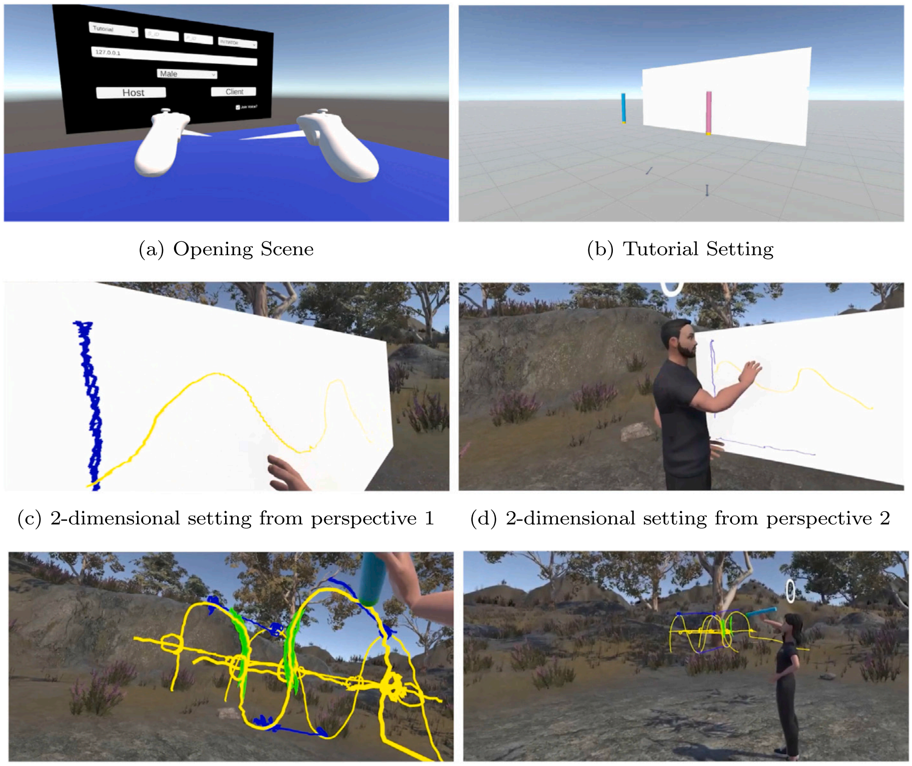
The literature discussed so far focuses on gestures and speech essential for communication during spatial dialogue in VR. However, as discussed in Section 2.1, interlocutors often complement gestures and speech with self-created props and prepared/improvised drawings. A separate branch of research in VR focuses on enabling authoring of digital content for collaboration within VR environments (Ververidis et al., 2022; Coelho et al., 2022). For instance, VRGit (Zhang et al., 2023) allows users to collaboratively design interior layouts using pre-made furniture models much like physical communication paradigms that rely on pre-made models in architecture. In addition, prior research has explored interactions to create 3D drawings in collaborative (He et al., 2020) and individual VR experiences (Dudley et al., 2018). The potential of 3D drawing tools has led to a detailed account of the technologies origin, benefits and challenges for supporting design and design collaboration in recent work by Arora et al. (2023). Specialized commercial VR applications, such as Arkio[1], have also been developed to cater to communication requirements around spatial concepts for the architecture, engineering and construction industry (Ververidis et al., 2022).

Despite the growing interest in using 3D drawings to enhance communication around spatial concepts, and the potential advantages of embodied and spatialized interactions for communication, prior work has largely neglected the influence of *creating* and *referring* to 2D or 3D drawings on interlocutors' actions and speech during spatial dialogue in VR. Our research aims to complement existing work on drawing supported communication practices (Snyder, 2013, 2014), and inform the design of future VR applications for spatial dialogue, by extending our knowledge of communication practices to virtual spaces that can leverage both 2D and 3D drawing modalities.

3. System description

We developed a custom multi-user VR application for the Meta Quest 2 to explore the interconnection of actions and speech that interlocutors employ when supporting spatial dialogue with 2D or 3D drawings

[1] https://www.arkio.is/



(a) Opening Scene
(b) Tutorial Setting
(c) 2-dimensional setting from perspective 1
(d) 2-dimensional setting from perspective 2
(e) 3-dimensional setting from perspective 1
(f) 3-dimensional setting from perspective 2

Fig. 1. Screenshots from our VR application. Figure a) shows the opening scene when users first open the application, b) shows the tutorial setting which includes both 2D drawing tool (pink pen and whiteboard) and 3D drawing tools (blue pen), c) & d) Perspective of two interlocutors when using the 2D drawing tools for communication in the VR application, and e & f) Perspective of two interlocutors when using the 3D drawing tools for communication in the VR application.

in VR. The VR application was made using OpenXR in Unity3D, along with Unity netcode[2] and vivox[3] to enable multi-user capabilities and proximity-based voice chat, respectively.

On starting the application, users are individually placed in a virtual environment. The environment is an empty space that includes only a horizontal plane that serves as the ground, and a user interface (UI) panel (Fig. 1(a)). Users can move within the virtual space by either moving in physical space, using the right controller to teleport to a desired location, or by using the left controller to smoothly transition across the space. The UI provides a visual interface for the researcher to set experimental parameters remotely via the computer tethered to the VR headset. These parameters include PARTICIPANT ID, SESSION ID, ROLE (determining whether this user initiates the dialogue), and SETTING (2D, 3D or Tutorial). The user only interacts with the UI panel to select a desired avatar (with Male or Female features), and to either host a collaborative session that additional users can join or join an existing session.

Depending on the SETTING selected, users will be placed in different virtual settings with access to various virtual tools for drawing. We first describe the various virtual tools in our VR application and then detail the virtual settings that provide access to these tools in the following sections.

3.1. Virtual tools

The complete set of tools includes two differently coloured virtual pens, and a virtual whiteboard (Fig. 1(b)). Users can interact with any available virtual tool by first 'grabbing' the tool using the grip button on the VR controllers. The tool can then be moved or used to perform actions (such as drawing, toggling eraser, erasing entire whiteboard, or changing the pen colour) using the VR controller that is currently 'grabbing' the tool.

Each virtual pen consists of a body and a tip. The colour of the body is used to differentiate between the pen that enables drawing on the virtual whiteboard (dubbed 2D pen) or drawing in 3D space via 3D lines/tubes (dubbed 3D pen). The tip colour represents the drawing colour, and both pens allow the user to switch between the colours red, green, blue, and yellow. In addition, both pens enable the user to toggle on/off an eraser mode used to remove drawings from either the whiteboard or 3D space.

The 2D and 3D pens were designed to have similar interactions. Both pens used the same VR controller buttons to change colour and toggle the eraser functions. Additionally, drawing with both pens required the user to hold down the trigger button on their VR controller. This method allowed users to control when to start and stop 3D drawings and also minimized accidental drawings when using the 2D pen with the whiteboard. Additionally, as the whiteboard presented limited space for drawing when compared to the entire virtual space, we provided users with the ability to quickly erase the entire whiteboard by using their VR controllers. Implementation of the drawing interactions was adapted from online video tutorials[4,5] and we include the Unity drawing-related scripts in our supplementary material.

As the focus of this paper is **primarily on spatial dialogue behaviours (verbal and non-verbal)** when using 2D and 3D illustrative tools in VR, we limit drawing capabilities that could introduce interactions requiring additional gestures unrelated to the communicative

2 https://unity.com/products/netcode
3 https://unity.com/products/vivox-voice-chat
4 2D: https://www.youtube.com/watch?v=sHE5ubsP-E8
5 3D: https://www.youtube.com/watch?v=UyGSOju9aWw&list=PLK56usTosKRe2wXchSs4FUP3uSSGIaHGA&index=21&t=435s

task. This led us to employ basic drawing interactions without advanced functions, such as translating and modifying drawings (available in commercial applications like Google's Tilt Brush) as these introduce the need for new controls and interactive gestures, which may influence non-verbal behaviours during spatial dialogue. Similarly, hand-based interactions were considered but were not used as these require users to learn and use hand gestures unrelated to communication to interact with our VR tools and environment. Additionally, the use of common hand-based gestures, such as pointing and pinching for object selection in VR, could deter interlocutors from performing deictic (pointing) gestures during communication to avoid unintended interactions — known as the 'Midas Touch' problem (Wu and Wang, 2016; Jacob, 1995). Finally, unlike VR controllers and hand-gesture recognition capabilities, VR stylus controllers are stand-alone equipment that must be procured independently. Challenges in obtaining stylus controllers at the time of our study prevented us from implementing and testing this interaction method. However, VR stylus controllers have the potential to provide more familiar drawing interactions, but may require additional controllers or hand-based interactions to enable non-drawing actions, such as grabbing the whiteboard or teleporting in the VR environment.

3.2. Virtual settings

The selected SETTING also determines the virtual setting in which users will collaborate. There are three different settings in our application; a) tutorial setting, b) 2D setting, and c) 3D setting.

Tutorial setting. This setting places users, as their chosen avatar, in an empty virtual space on a horizontal plane, akin to the opening virtual environment. Unlike the opening setting, the tutorial setting is a multi-user space and enables multiple users to inhabit the same virtual space. This space does not include a UI panel but is the only setting that contains all three virtual tools described in the previous section (2D pen, 3D pen, and virtual whiteboard). This setting is designed to introduce users to the multi-user virtual environment and the drawing tools available in our VR application.

2D setting. As spatial context is relevant to interlocutors engaged in spatial dialogue (Coventry et al., 2009), including in VR (Smith and Neff, 2018), users are placed in a virtual terrain modeled to resemble a real-world geological environment, providing context for the structural geology-related dialogue in the experimental task (Fig. 1). This enables us to explore the connections between the gestures and language that interlocutors use around environmental and drawing-related frames of reference, as mediated by the drawing tools available (2D or 3D) during spatial dialogue.

In our 2D setting, users are provided with one 2D pen that can be used to draw on a provided virtual whiteboard. In addition, the whiteboard can be placed anywhere in the virtual space using basic grab and release interactions also used for the 2D and 3D pens and is not affected by gravity. Capabilities to move the whiteboard were provided to alleviate discrepancies between 2D and 3D settings in situating drawings within the virtual environment for communicative purposes. Any user can interact with any virtual tool that is not currently in use by another user.

3D setting. The 3D setting also places users in the same virtual geological terrain used in the 2D setting. However, they are only given access to a single 3D pen for creating 3D drawings anywhere in the virtual space. Drawings made in 3D space are immovable, and like the 2D setting, any user can interact with the 3D pen provided if it is not currently in use by another user.

4. Method

In this paper, we aim to better understand the interconnection of actions and speech that interlocutors employ when supporting spatial dialogue with 2D or 3D drawings in VR. As such, we conducted a within-subject user study where a pair of participants (dyad) were tasked with discussing a given topic that involved spatial concepts while supported with 2D or 3D drawing tools in a collaborative VR application (see Section 3). All our methods received approval from our institutions ethical review panel.

4.1. Experimental design

We investigate the effects of 2D and 3D drawing tools on spatial dialogue by observing, and analyzing video and audio recordings of participating dyads engaged in discussions involving spatial concepts while manipulating access to supporting drawing tools. Specifically, our experiment involved 2 conditions with a single independent variable related to DRAWING DIMENSION (2D or 3D) that determined the capabilities of the drawing tool available to participants within the VR application. Each dyad participated in both conditions with a different task assigned to each condition. Each task required dyads to discuss a prescribed topic involving spatial concepts with the help of the assigned drawing tools. We chose topics related to structural geology as drivers for the discussion in our study. This decision was based on the importance of spatial relationships and concepts within structural geology, as well as our team's experience with the chosen topics — which includes a senior academic specializing in structural geology.

We recorded participants' voices and first-person perspective views while using our VR application during each **conversation**; which refers to the dialogue during a single condition employing 2D or 3D tools to support discussion on the given topic in one experimental session. The recordings are used to better understand the effects of DRAWING DIMENSION on interlocutor's behaviour (gestures, drawing behaviours, and spatial organizations) and speech. We further logged completion times (when both participants agreed to stop the dialogue) and total time spent on drawing (logged by our VR application) to understand performance differences between 2D and 3D drawing tool use. We also collected post-experiment measures including; 1) the NASA-TLX questionnaire to investigate task-load between the conditions, 2) subjective opinion measures on communication effectiveness using custom Likert-scale questionnaires (detailed in Appendix A.1), and 3) post-test questionnaires to gain preliminary insights into interlocutors understanding of topic content. Finally, we recorded an interview with the dyad focusing on their experience using the different drawing tools to discuss spatial concepts.

4.2. Participants

Twenty participants grouped into 10 dyads (Male: 12, Female: 7, Non-Binary: 1, Aged M: 30.8, SD: 6.95) took part in our experiment. We employed convenience sampling to recruit participants, and interested individuals were asked to sign up with a known individual to control for effects related to communicating with strangers (Duronto et al., 2005). Participation was voluntary and not reimbursed.

4.3. Apparatus

The experiment took place in a controlled laboratory space that was divided into two sections. Each section consisted of a table with a desktop computer (Processor: 12th gen Intel(R) Core(TM) i7-12700, 4.90GHz, Core Count: 8 + 4; Memory: 32768MB 4400MHz DDR5; Graphics Processing Unit: NVIDIA 3090 RTX) and a 2×1.8 meter subsection that was cleared of all objects. The sections were spaced 2 meters apart from each other, and each subsection functioned as the designated usage *zone* for our VR application. As our VR application included a large outdoor terrain environment with sizeable texture files (see Section 3.2), we were required to tether our Meta Quest 2 head-mounted displays to the desktop computers in order to run our application on the headsets through the computers without performance issues.

Table 1
Summary of participant demographic data.

Gender	Age	Exp. HMD [0–5]	Exp. 3D games [0–5]	Exp. 3D thinking & reasoning [0–5]	Exp. teaching & mentoring [0–5]	Exp. field trips [0–5]	Exp. structural geology [0–5]
Male: 12 Female: 7 Non-binary: 1	Mean: 30.85 SD: 6.95	Mean: 3.2 SD: 1.5	Mean: 3.4 SD: 1.5	Mean: 4.2 SD: 0.6	Mean: 3.9 SD: 1.0	Mean: 3.2 SD: 1.3	Mean: 2.0 SD: 1.0

4.4. Procedure

Participating dyads first read a plain-language statement detailing our study and were asked to provide written consent. We then collected demographic details related to age and gender. We also collected 5-point Likert-scale responses to gather data on participant's experience with head-mounted displays (augmented, mixed, or virtual reality HMDs), 3D multiplayer games, 3D thinking and reasoning, teaching and mentoring, field trips to nature, and structural geology. Summary of participant demographic data is presented in Table 1.

Participants were then led to separate VR *zones* and were asked to put on the VR headset and hold the VR controllers. The researcher also assisted participants in connecting and putting on a set of headphones with an in-built microphone. The headphones served the purpose of blocking out external noise, communicating via our application's spatial voice chat feature, and for audio recording the conversation during the experimental session. Prior to every session, the researcher cleaned all hardware used by participants and ensured that each VR 'zone' was set up correctly (via the Meta Quest 2 Guardian function[6]).

After ensuring that both participants were comfortable with the placement of their headsets, a researcher started the custom VR application for both participants. Participants were then instructed to select an avatar from the UI in the opening VR environment (see Section 3). One participant was then asked to host a session with SETTING: TUTORIAL, and the other participant was asked to join the host's session[7]. Once both participants were inside the tutorial setting (see Section 3.2) and could see each other's avatars, the researcher explained the basic controls for VR movement, object interactions and use of the different custom virtual tools available for drawing. Participants were asked to perform the different interactions to follow along with the researcher's explanation. Additionally, as our system was developed using Unity netcode which interpolates virtual object positions between clients, drawings made with very fast and small movements may not be accurately depicted to observing users. As such, we instructed participants not to use excessively fast or small movements when drawing, and to ask for clarification from the person drawing if they noticed any discrepancies between language and the referent drawing. Finally, participants were given a maximum of 10 min to use the virtual tools and clarify any doubts related to tool use with the researcher.

Participants were then randomly assigned the role of an *Initiator* or a *Discussant*, which were maintained across the two conditions in one experimental session. In both conditions, the *Initiator* first watches an approximately 5 min long video lecture on a given topic in structural geology involving spatial concepts. Each condition was prescribed a different video lecture that was adapted from a longer lecture delivered to undergraduate geology students at our university (The specific videos were on the topics of 'Dip and Strike' and the geological structure of 'Folds'). Both participants would then put on the VR headsets and begin the experimental trials, which placed them in the virtual terrain modeled after a geological environment (see Section 3.2). Participants were not informed of, or exposed to, the virtual terrain prior to the experimental trials to enable us to observe if, and how, participants spontaneously used the virtual terrain to support the experimental task. The *Initiator* would then have 7 min and 30 s[8] to explain the concepts presented in the video to the *Discussant* in VR with access to drawing tools determined by the experimental condition. The *Discussant* was allowed to ask questions and add to the conversation during the *Initiator*'s explanation. After the *Initiator* finished their explanation, the *Discussant* was given 7.5 min to re-iterate their understanding of the topic to the *Initiator* with access to the same drawing tools. The *Initiator* was similarly allowed to ask questions and add to the conversation without explicitly correcting the *Discussant* on misunderstood points. Both *Initiator* and *Discussant* were only tasked with explaining the given content, and had no requirements for the number of drawings to create in order to complete the experimental task. If participants completed their explanation within the allotted time, they were asked to signal the researcher to move on to the next step.

As we employed a within-subject approach, we counterbalanced our experiment based on our DRAWING DIMENSION variable to ensure that we accounted for order effects. Additionally, as we do not compare across the different topics, the order of topics remains the same in all experimental sessions. This was to ensure that we had an even distribution of participants discussing the first and second topics using the 2D and 3D drawing tools. Participants also maintained their roles across both conditions (2D and 3D) in one experimental session to enable meaningful comparisons of their experiences between the two conditions.

After each condition, participants were asked to set aside the VR equipment and fill out a NASA-TLX questionnaire to measure workload, followed by a custom Likert-scale survey to measure participant's opinions on communication effectiveness while supported by the available VR drawing tools in that condition. The Likert-scale survey consisted of 5 ratings (from 'Strongly Disagree' to 'Strongly Agree') and was designed to measure participant's subjective opinions on the usefulness of the drawing tool, and the usefulness of the virtual environment when communicating. In addition, the Likert-scale measured the *Initiator's* opinions on their understanding of the prescribed topic content, success in communicating their understanding to the *Discussant*, and their confidence that the *Discussant* understood their explanation. Similarly, the Likert-scale also measured the *Discussant's* opinions on their understanding of the *Initiator's* explanation, and their success in re-iterating the content to the *Initiator*. While our study was not focused on content understanding/learning outcomes, we also conducted post-test questionnaires to gather preliminary insights on the effects of 2D and 3D drawings on user understanding of spatial concepts related to structural geology. At the end of the experimental session, we conducted a *brief* semi-structured interview to gather insights on the user experience and opinions using the different drawing tools with DRAWING DIMENSION: 2D OR 3D. Details on the custom Likert-scale survey, the post-test questionnaires, and the semi-structured interview are presented in our Appendix A.

[6] https://www.meta.com/en-gb/help/quest/articles/in-vr-experiences/oculus-features/boundary/

[7] The network address needed to join the session was set by the researcher

[8] Two and a half minutes longer than the video content to account for unfamiliarity with the topic

5. Analysis

5.1. Qualitative data

Our qualitative data include 7 h and 23 min of audio and video recordings of dyads engaged in conversation while using our VR application. This includes 2 recordings of each participant's perspective for each condition of our experiment with 10 experimental sessions, resulting in a total of 40 recordings (2 participant recordings x 2 conditions x 10 sessions). We also collected 1 h and 56 min of audio recordings in our post-experiment interviews with the dyads.

5.1.1. Conversation recordings

We first generated transcripts for each recording using Microsoft OneDrive. We then grouped recordings and transcripts of the same conversation from the perspective of 2 participating interlocutors to be examined together. We used open coding to identify and code different sections of the video recordings that were relevant to spatial dialogue. Coding was performed on the conversation recording transcripts in NVivo while simultaneously viewing the video. Specifically, we coded for *user actions that mediated speech*, including the use of gestures, deictic expressions, interlocutors' changes in spatial organization within the VR environment (to then reference objects from a different viewpoint, for example), and use of drawings to support verbal communication. One researcher viewed videos of each conversation, alongside the generated transcripts, to check for, and correct, any errors during transcription. During this first viewing, the researcher also developed initial codes. A second viewing of the videos was then carried out by the same researcher to further develop the codes and ensure consistency of the codes that were developed during the first viewing.

Additionally, during the second viewing of the videos, we further refine and categorize our codes. We also consider the rich body of existing literature around the categorization of different gestures during speech (McNeill, 1992; Ekman and Friesen, 1969; Efron, 1941) in the process of refining our codes related to gestures. Specifically, we classify the gestures observed in videos based on Ekman and Friesen (1969) categorizations of the kinesic movements of *illustrators* — which Ekman and Friesen (1969) define as 'movements that are directly tied to speech, serving to illustrate what is being said verbally'.

A second researcher was invited to independently code the conversation recordings for sections relevant to spatial dialogue. 20% of the recordings (2 sessions, inclusive of 2D and 3D conditions) were assessed for inter-rater reliability through Nvivo coding comparison analysis. A total of 460 coding references were made by both coders across the 2 sessions — out of which 193 coding references were reported as mismatches associated with codes having a < 0.4 Kappa coefficient. The coders collectively reviewed the 193 mismatched references over 4 one-hour long online video meeting sessions. The majority (n = 176) of these codes were mismatched in the analysis due to a lack of character overlap. This resulted from the combined use of transcribed audio with video recordings (needed for gestures), whereby one coder may have selected the precise timestamp where a gesture occurred, while the other coded the related transcribed speech. Other codes were quickly resolved where they were identified as errors. Coders only disagreed on 17 coding references (3.7%) of the total 460 coding references made by both coders in these two sessions, resulting in a high inter-rater agreement of 96.3%.

5.1.2. Epistemic network analysis

We employ epistemic network analysis (ENA) using the ENA Web Tool (Marquart et al., 2018) on our coded data. ENA enables us to identify, quantify, and dynamically visualize the structure of connections among the different elements in our coded data (Shaffer et al., 2016; Syiem and Türkay, 2024) for each *conversation* (unit of analysis). The relationships between the elements are measured by quantifying the co-occurrence of those elements in discourse (Shaffer et al., 2009). ENA does this by creating an adjacency matrix of the frequency of co-occurring elements, transforming the matrix into a vector, and normalizing the vector by its length. These normalized vector values represent the relative strength of connections (co-occurring elements) in each conversation (which can also be aggregated over a number of conversations (Shaffer et al., 2016)). ENA then projects this representation onto a 2-dimensional space using singular value decomposition (SVD) or means-rotation (MR) along the x-axis [9], along with SVD on the y-axis (Bowman et al., 2021; Tan et al., 2024; Rolim et al., 2019)[10]. The decomposed values can further be used to statistically test for differences between groups in the data using t-tests or the Mann-Whitley U test (Tan et al., 2024).

ENA was originally developed to model connections between different discourse elements (vertices/nodes), with the underlying assumption that the connections between different nodes (edges) are more important than the presence of the node itself (Bowman et al., 2021). ENA has been used to model discourse networks based on cognitive theories (Bowman et al., 2021) (see examples (Arastoopour Irgens et al., 2015; Thompson et al., 2021; Bressler et al., 2019), which include analysis of spatial dialogue related utterances in a VR game (Uz-Bilgin et al., 2020)), but it has seen applications in various fields (Elmoazen et al., 2022) including social networks (Nash and Shaffer, 2013), social gaze coordination (Andrist et al., 2015), and surgeon's communication during operative procedures (D'Angelo et al., 2020), among others.

In our study, we use ENA to model the weighted network of *actions used to mediate speech* that dyads perform to support spatial dialogue in VR while using 2D or 3D drawing tools. We model individual networks *for each conversation* recorded during our experiment. We employ means rotation (MR) for dimensional reduction along the x-axis as our data consists exactly of 2 groups (Drawing Dimension: 2D and Drawing Dimension: 3D) (Bowman et al., 2021). We then use the centroid of individual networks, which summarizes the structure of connections within an individual network, to aggregate and compare networks associated with 2D and 3D drawing conditions (Shaffer et al., 2016). Comparison of ENA networks can be achieved by *visually* inspecting different networks (aggregated or individual) to identify differences in connections. *The ENA Web Tool also enables the visual representation of the difference between two networks by subtracting the corresponding edge weights of the individual networks. The resulting network exhibits edge colours relating to the network with stronger co-occurrence for that edge, and edge weights signifying the magnitude of the difference in co-occurrence between the two networks.* In addition, ENA enables the use of *statistical* methods to analyse the difference between the two networks. Details related to statistical comparisons between networks and the mathematical foundations of ENA are beyond the scope of this paper and can be found in prior work (ENA Web Tool (Shaffer et al., 2016), ENA with R (Tan et al., 2024), and mathematical foundations of ENA (Bowman et al., 2021)).

We use the networks generated through ENA to better understand how interlocutors connect different communicative actions during drawing supported spatial dialogue in VR. As an example, 2D drawing capabilities may elicit more connections between actions of drawing to support speech and deictic movements to referents in the virtual environment, in order to form associations between the abstracted 2D drawings and the actual 3D object they represent. In contrast, 3D drawings may prompt more connections between drawing related actions and spatial organizations within the virtual environment, in order to anchor drawings meaningfully in the virtual environment.

5.1.3. Interview recordings

Interviews were kept brief to minimize respondent fatigue (Porter et al., 2004) as they were conducted after participants took part in two

[9] Means rotation is a data projection method that is uniquely developed for ENA. While beyond the scope of this paper, the mathematical concepts around means rotation are presented in Bowman et al. (2021) paper.

[10] An overview of ENA interpretations can be found in the ICQE21 Workshop: Advanced ENA Interpretations video: https://www.youtube.com/watch?v=_2LpAefX8KM

experimental conditions (and onboarding), and had completed multiple post-study questionnaires and surveys (detailed in Section 4.4). As such, interview data were primarily used to understand users' experiences, and to supplement our audio & video recording data of the experimental sessions. An average of 5.8 min of interview data was recorded per participant (116 min/20 participants). Given the short duration of the interviews, we opted for manual examination of the interview transcripts. This involved two researchers independently identifying participant quotes relevant to our research aims, and conducting multiple rounds of discussion (online and in-person) to arrive at a shared view of participant perceptions and user experiences. One researcher then cross-examined the identified quotes with findings from the analysis of the conversation recordings.

5.2. Quantitative data

Our within-subject experiment was designed to investigate the effects of Drawing Dimension on dependent measures, including completion time, total time on drawing, NASA-TLX measures, post-test scores, and Likert-scale surveys on perceived communication effectiveness (see Section 4.4). However, as participants were assigned a specific *Role* (Initiator or Discussant) during an experimental session (see Section 4.4) that could influence task workload, post-test scores & survey responses, we group related analyses based on both Drawing Dimension and *Role*. We employed paired sample t-tests or Wilcoxon signed-rank tests for statistical analysis of numerical data (completion time and total time on drawing) dependent on the normality of the data as determined by the Shapiro-Wilk's test. In addition, we conducted statistical tests using cumulative link mixed models (CLMMs) with a 'probit' link function (also known as ordered probit models) (Christensen, 2018) to assess effects of participant *Role* (Initiator or Discussant) and Drawing Dimension on ordinal measures (Liddell and Kruschke, 2018) (NASA-TLX measures, post-test scores, and Likert-scale survey responses) while accounting for random effects related to each participant. While the effects of *Role* are not a focus of this paper, we present the results of our analysis for completeness.

Finally, Likert-scale survey responses that were unique to each *Role*, were analysed for significant effects of Drawing Dimension using CLMMs while accounting for random effects related to each participant. The specific measures include Initiators subjective ratings on their understanding of the video content, their communication with the Discussant, and their perception of the Discussant's understanding of their explanation, and the Discussant's subjective ratings on their understanding of the Initiators explanation, and their re-iteration of the content delivered by the Initiator.

6. Findings

6.1. Qualitative data

Using the analysis procedure described in Section 5.1.1, we coded 870 instances of users employing actions that immediately, or in future, mediate speech (for instance, changing positions around an object to enable deictic words to be meaningful from the new point of view). A total of 14 codes were developed during our analysis; we detail each code below:

- ***Drawing Mediated Speech*:** Describes instances where interlocutors' speech depends on referents that are simultaneously being created through drawing. For example, the sentence from our first experimental session (S1) — "***that's*** the dipping direction of the plane, ***this*** new plane" while the speaker was in the process of drawing the 'new plane' was coded as *Drawing Mediated Speech.*
- ***Illustration Referencing Drawing*:** Describes instances where interlocutors' speech depends on referents that are simultaneously being created through drawing, while also relating the new drawing to a previously created drawing/illustration. An example of this can be seen in the following lines from S9: "We could plot ***this*** [referring to a previous drawing] around ***here*** [referring to the current drawing] if we count from the outside being 0 and the inside being … pointing directly down. And in 3D it would go like,…**that** [referring to current drawing] would represent **that** [referring to previous drawing] surface".
- ***Environment Situated Drawing*:** Describes instances where interlocutors discuss concepts with the help of drawings that are made in close spatial proximity to an object related to the concept in the virtual environment. This could be in 3D by drawing directly in close proximity to the object of interest, or in 2D where the users can move the whiteboard closer to objects of relevance to then draw on. For example in S10, the sentence "And then the other angle is …***where am I*** …***this*** angle", indicates an interlocutor trying to find the correct spatial position next to a relevant virtual object (***'where am I'***), to then being drawing ('***this*** angle').
- **Spatial Organization:** Describes instances where both participants in a conversation *collectively* reorient themselves in space around a drawing, each other, or a virtual object in the virtual environment. This can be prompted (coded as ***Prompted Spatial Organization***) or unprompted (coded as ***Unprompted Spatial Organization***). An example of a prompted instance was observed during S3 where the speaker is prompting their collaborator to move: 'so if you move … behind me'. Unprompted instances are typically not immediately observed in speech (speech is mediated after the change), but were observed in the video recording. For example, a speaker moves to a new blank section of the whiteboard to create and describe a new drawing and the collaborator follows.
- ***Speaker Pose Change*:** Describes instances where the speaker changes their viewpoint (orientation and/or position) to leverage 3D space. This could be to draw a 3D drawing, or to discuss or question an already created drawing from a different viewpoint. For instance, in S9 our speaker projects a 2- dimensional point into a line to illustrate the 3-dimensional structure the point represents while saying: "And then I guess in 3D, these form lines … like this through the folds …. ".
- **Kinesic Illustrators:** Describes instances where users employ different Kinesic illustrators to complement their speech during the conversations in VR. The different kinesic illustrators are classified based on Ekman and Friesen (1969) categorizations which include;
  - ***Baton***: Movements used to emphasize particular words. For example, making a sharp, short and rapid hand movement when saying 'No, this is *much larger* than the previous crystal we saw' to emphasize the words 'much larger'.
  - ***Ideograph***: Movements that illustrate an idea or abstract concept. For example, a speaker opening their arms starting palms down at the closed arm position to palms up at the open arm position to indicate the space around them when saying 'Oxygen is *everywhere*'.
  - ***Kinetograph***: Movements representing a kinetic behaviour. For example, speaker moves hands at an angle from a higher to a lower position while saying 'The steepest trajectory can be determined by how water would *run down the slope*'.
  - ***Pictograph***: Movements illustrating a real object. For example, a speaker moving their hand in an upside down 'J' shape when making a description — 'The man was carrying a cane that looked *something like this*'.
  - ***Spatial Gesture***: Movements depicting a spatial relationship (direction, distance, orientation, etc.). For example, placing their hands apart from one another and then moving them further apart to signify relative distance when saying 'if the closest gas station was *this* far, the next gas station is about *three times* the distance'.
  - **Deictic Movement**: Pointing gestures to a present object. These gestures are coded as; ***Deictic Drawing Reference*** representing pointing gestures towards an existing drawing, ***Deictic Environment Reference*** representing pointing gestures towards the virtual environment, and ***Deictic Tracing***

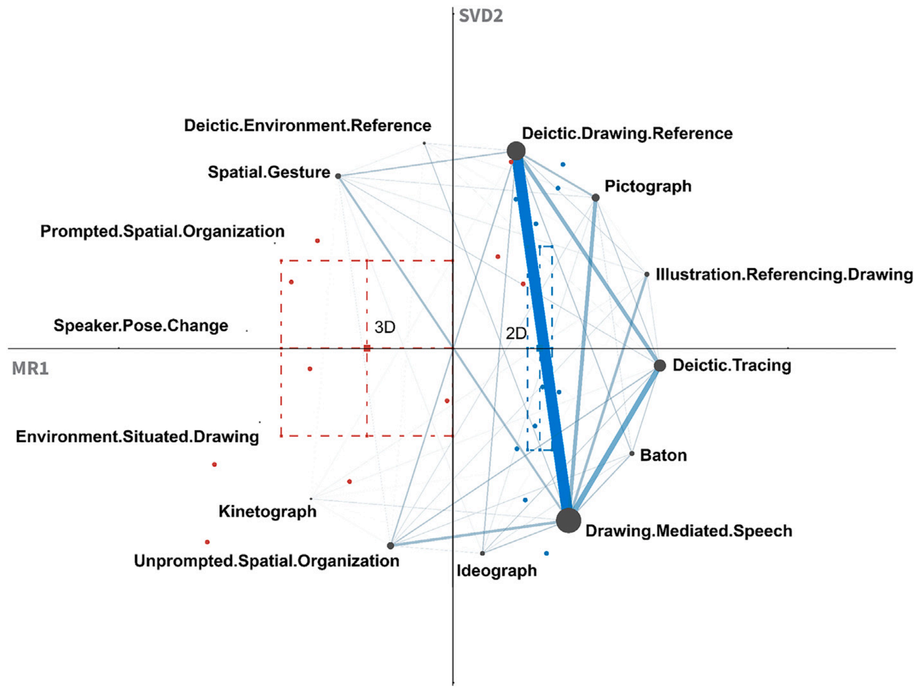


Fig. 2. The ENA network for interlocutors in the 2D VR drawing supported spatial dialogue condition. The network illustrates the co-occurrence frequency of our codes (represented by black nodes) via a weighted edge (blue line connecting two black nodes). The size of a code (black) node is proportional to the frequency of that code in *all experimental sessions* in the 2D condition. The weight of an edge connecting two code nodes is proportional to how frequent the two codes appeared together within *all experimental sessions* in the 2D condition. Small coloured nodes represent an ENA network centroid for a *single experimental session* in the 2D (blue nodes) or 3D (red nodes) condition. The position of these small coloured nodes relative to different edges and code nodes indicate which codes (communicative actions) were more prominent in those individual sessions. In our 2D condition, we observe the strongest connection between communicative actions of *Drawing Mediated Speech* and *Deictic Drawing Reference*, with relatively strong connections between *Drawing Mediated Speech* and *Deictic Tracing, Illustration Referencing Drawing* and *Pictograph*. No edges connected from any code to *Prompted Spatial Organization*, *Speaker Pose Change*, and *Environment Situated Drawing*.

(Schueler and Wesslein, 2022) representing movements that combine *pictographs* with *deictic drawing reference*, which is a special case of kinesic illustrator that emerged during our analysis.

Rich, illustrative worked examples that further detail kinesic illustrators can be found in McNeill (1992) work for the interested reader.

We then created an ENA model to compare the weighted connections of actions between 2D and 3D drawing supported spatial dialogue in VR. Our unit of analysis for ENA refers to a single *conversation*, on a single prescribed *topic* supported by a specific drawing tool (Drawing Dimension:2D, 3D). Additionally, we consider the connections of our developed codes within the whole conversation.

Figs. 2 and 3 show the network for conditions with Drawing Dimension:2D and Drawing Dimension:3D, respectively. Each black node represents a code, and each coloured node represents a centroid of a network for an individual conversation supported by Drawing Dimension:2D (blue) and Drawing Dimension:3D (red)[11]. Connecting edges between two codes signify co-occurrence, and the thickness of the edge indicates the strength of the connection (Table A.8 in our appendix presents the connection strength values of all edges as determined by the ENA tool). Finally, the coloured square points and the dotted bounding boxes represent the associated mean and confidence intervals (along the x- and y-axis) for each condition.

The networks for both 2D and 3D conditions indicate strong connections between *Drawing Mediated Speech* and *Deictic Drawing Reference*. This suggests that interlocutors' not only adapted their speech by creating new referents via drawing, but frequently had to refer to already created drawings. Both networks also show relatively strong connections between *Deictic Tracing*, *Deictic Drawing Reference*, and *Drawing Mediated Speech*, indicating actions that reinforce communication via drawing with pointing, as well as tracing, gestures. An expected distinction between the networks, arising from the additional dimensionality afforded by the 3D drawing tool (Arora et al., 2023), is the absence of connections related to *Speaker Pose Change*, *Prompted Spatial Organization*, and *Environment Situated Drawing* in the 2D conditions. This suggests that interlocutors' *recognized and used* the unique spatial affordances of 3D drawing tools in VR to support verbal communication of spatial concepts.

As the coded elements (black nodes) in an ENA network are consistently positioned between plots, we can visualize networks representing the differences between connections in the Drawing Dimension: 2D and Drawing Dimension: 3D conditions (Shaffer et al., 2016). Fig. 4 shows the difference between the connections of coded elements in Drawing Dimension:2D and Drawing Dimension:3D. Blue edges indicate stronger connections in the 2D condition, and red edges indicate stronger connections in the 3D condition. Stronger connections between *Drawing Mediated Speech*, *Deictic Drawing Reference*, and *Pictograph* in the 2D condition, suggest more frequent use of pointing at created drawings, along with mid-air gestural illustrations (without a referent) to reinforce communication using drawings. Additionally, there were stronger connections between *Drawing Mediated Speech* and *Illustration Referencing Drawing* in the 2D condition, indicating that users created more additional drawings to relate to existing drawings while using 2D drawing tools when compared to 3D drawing tool use.

In contrast, we found stronger connections between *Speaker Pose Change* and *Deictic Drawing Reference*, and *Speaker Pose Change* and

[11] We chose the 'Unit Circle - Equally Spaced' option on the ENA Web Tool for better visibility of the nodes

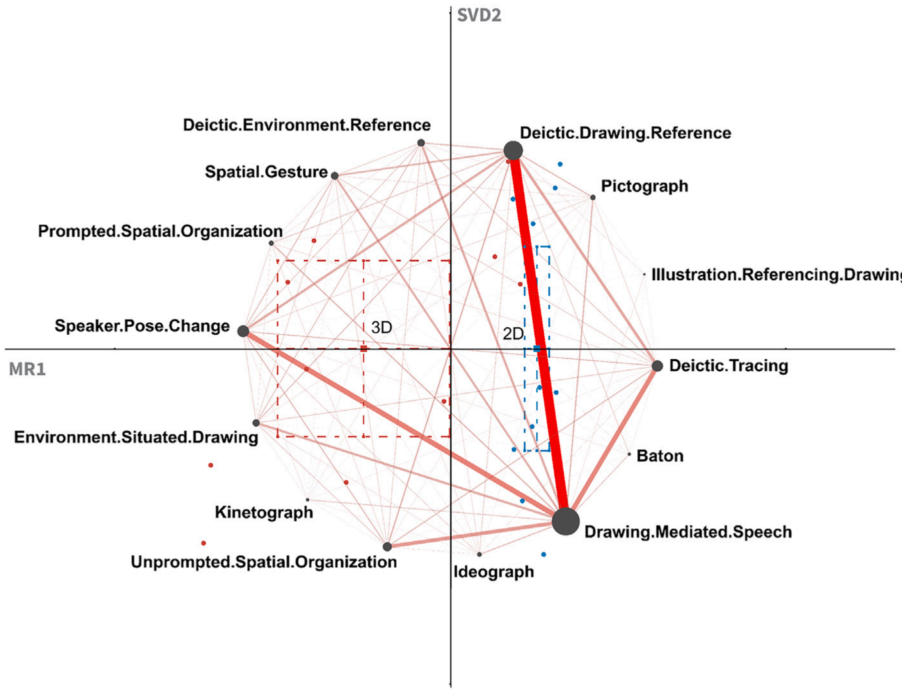


Fig. 3. The ENA network for interlocutors in the 3D VR drawing supported spatial dialogue condition. The network illustrates the co-occurrence frequency of our codes (represented by black nodes) via a weighted edge (red line connecting two black nodes). The size of a code (black) node is proportional to the frequency of that code in *all experimental sessions* in the 3D condition. The weight of an edge connecting two code nodes is proportional to how frequent the two codes appeared together within *all experimental sessions* in the 3D condition. Small coloured nodes represent an ENA network centroid for a *single experimental session* in the 2D (blue nodes) or 3D (red nodes) condition. The position of these small coloured nodes relative to different edges and code nodes indicate which codes (communicative actions) were more prominent in those individual sessions. In our 3D condition, we observe the strongest connection between communicative actions of *Drawing Mediated Speech* and *Deictic Drawing Reference*, with relatively strong connections between *Drawing Mediated Speech* and *Deictic Tracing*, *Speaker Pose Change*, and *Unprompted Spatial Organization*.

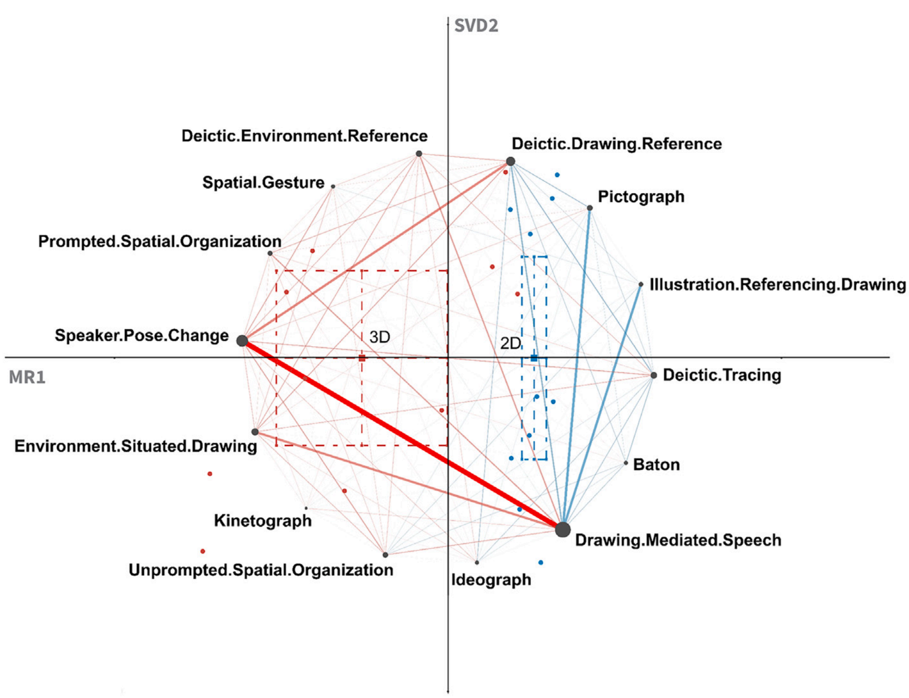


Fig. 4. An ENA network that visually depicts the differences in co-occurrence frequency of coded elements (communicative actions) between our 2D and 3D conditions (process detailed in Section 5.1.2). This network exhibits blue edges for a pair of codes that occurred together more frequently in the 2D condition, and red edges if co-occurrence of the corresponding codes were more frequent in the 3D condition. We observe that the connections between *Speaker Pose Change* and *Deictic Drawing Reference*, between *Speaker Pose Change* and *Drawing Mediated Speech*, and between *Environment Situated Drawing* and *Drawing Mediated Speech*, were stronger in the 3D condition. Whereas, connections between *Pictograph* and *Drawing Mediated Speech*, and between *Illustration Referencing Drawing* and *Drawing Mediated Speech* were stronger in the 2D condition.

Table 2
Mean and Standard deviation (within parenthesis) of completion times and total time spent on drawings (drawing time) grouped by Drawing Dimension. Dyads performed similarly in both conditions related to 2D and 3D drawing tools in VR.

Drawing dimension	N	Completion time (seconds)	Drawing time (seconds)
2D	20	681.5 (184.1)	45.6 (30.1)
3D	20	643.3 (189.8)	44.7 (29.4)

Drawing Mediated Speech in the 3D condition. This indicates an association between changing 3D perspective views and pointing at created drawings from different angles to support communication using 3D drawing tools. Additionally, we found stronger connections between *Drawing Mediated Speech* and *Deictic Environment Reference*, and *Environment Situated Drawing* in the 3D condition when compared to 2D. This suggests that conversations supported by the 3D drawing tool observed more user behaviour that referred to, and anchored drawings on, the virtual environment. Finally, we also found connections between *Prompted Spatial Organization* and *Drawing Mediated Speech* in the 3D condition, indicating that users more frequently had to prompt their collaborators to spatially reorient themselves in the 3D condition than in the 2D condition.

The differences found by visually analysing the networks for conditions related to *Drawing Dimension: 2D* and *Drawing Dimension: 3D* were also tested for statistical differences using a two-sample t-test (assuming unequal variance) along the x-axis (mean rotated value, MR1). We found a significant difference in coding co-occurrence between *Drawing Dimension: 2D* and *Drawing Dimension: 3D* ($mean = -0.2182$, $SD = 0.3026$; $t(9.3665) = -4.5157$, $p = 0.0013$, $Cohen's\ d = 2.0195$). This suggests that interlocutors used significantly different connections of actions using 2D and 3D drawing tools to communicate spatial concepts in VR.

6.2. Quantitative data

6.2.1. Performance time measures

Table 2 details the summary statistics for the task completion times and total time spent on drawing (drawing time) grouped by Drawing Dimension. The data indicate that participants in both conditions spent similar amounts of time conversing about the task topic in VR. Additionally, participants also spent similar amounts of time drawing in both 2D and 3D conditions. A Wilcoxon signed-rank test also revealed no statistically significant effects of Drawing Dimension on *completion time* ($W = 120$, $p = 0.596$, $r = 0.125$) and on *drawing time* ($W = 111$, $p = 0.841$, $r = 0.050$), i.e., we found no evidence to suggest that 2D and 3D drawing tools in VR influenced the amount of time spent on drawings and in the total time spent discussing spatial concepts.

6.2.2. Subjective workload

Table 3 provides the mean workload and standard deviation for all sub-scales and the overall workload of the NASA-TLX form as reported by participants, grouped by Drawing Dimension and participant *Role*.

Table 4
Mean and Standard deviation (within parenthesis) of the post-test scores grouped by Drawing Dimension and *Role*. As expected, post-test scores for participants with *Role: Initiator* were larger than participants with *Role: Discussant.*

Drawing dimension	Role	N	Post-test score [0-12]
2D	Initiator	10	8.2 (2.8)
3D	Initiator	10	8.1 (3.1)
2D	Discussant	10	3.7 (3.9)
3D	Discussant	10	5.0 (4.0)

All sub-scales, with the exception of *physical demand* indicate similar values when grouped by *Role*. Statistical testing using CLMM revealed no evidence that Drawing Dimension ($b = -0.282$, $z = -0.602$, $p = 0.547$), *Role* ($b = -0.450$, $z = 0.408$, $p = 0.683$), or their interaction ($b = 0.204$, $z = 0.306$, $p = 0.759$) were significant predictors of overall workload. Additionally, Drawing Dimension, *Role*, and their interaction were not found to be significant predictors of any of the individual NASA-TLX sub-scale measures.

6.2.3. Post-test scores

Table 4 provides the summary statistics for the post-test scores achieved by our participants, grouped by Drawing Dimension and participant *Role*. The post-test scores for participants in the *Role: Initiator* were much larger than those in *Role: Discussant*. Statistical tests using CLMM indicate that *Role* is a significant predictor of post test scores ($b = 1.655$, $z = 2.766$, $p = 0.005$). This is expected as participants in the *Role: Initiator* group had additional content relevant to the post-test questionnaire in the form of the video lecture (see Section 4.4). The data also show similar performance between 2D and 3D Drawing Dimensions within each *Role*, with a slightly higher score for *Discussants* in the 3D condition when compared to 2D. However, our analysis using CLMM did not indicate that Drawing Dimension ($b = 0.512$, $z = 1.091$, $p = 0.274$) or the interaction between *Drawing Dimension* and *Role* ($b = -0.525$, $z = -0.795$, $p = 0.426$) were significant predictors of post-test scores.

6.2.4. Likert-scale responses

Table 5 presents the summary statistics for responses to both common and unique survey questions in relation to the different participant *Roles*, grouped by Drawing Dimension and participant *Role*. The data for participants in both the *Role: Initiator* and *Role: Discussant* groups suggest a more favourable opinion towards the 3D drawing tools across all subjective measures when compared to the 2D condition (measures are detailed in Section 4.4, caption for Table 5, and the complete surveys used are presented in the Appendix A.1). Statistical tests using CLMM revealed that Drawing Dimension ($b = 0.868$, $z = 1.391$, $p = 0.164$), *Role* ($b = -0.465$, $z = -0.681$, $p = 0.496$), and their interactions ($b = 0.458$, $z = 0.561$, $p = 0.575$) were not significant predictors of usefulness of the drawing tool in supporting conversation (I.D.Usefulness DT). Similarly, Drawing Dimension ($b = 0.898$, $z = 1.664$, $p = 0.096$), *Role* ($b = -0.01$, $z = -0.024$, $p = 0.981$), and their interactions ($b = -0.494$, $z = -0.682$, $p = 0.495$) were found to not significantly predict usefulness of the virtual environment in supporting conversation

Table 3
Mean and Standard deviation (within parenthesis) of the NASA-TLX scores for each sub-scale grouped by Drawing Dimension and Participant *Role*. All sub-scales indicate that subjective workload was similar across both *Role* and *Drawing Dimension*. Physical demand sub-scale indicates a lesser load for *Role: Initiator* than *Role: Discussant*.

Drawing dimension	Role	N	Mental demand	Physical demand	Temporal demand	Performance[a]	Effort	Frustration	Overall
2D	Initiator	10	55.7 (21.2)	19.7 (13.0)	34.7 (24.4)	45.7 (30.5)	53.7 (14.7)	38.7 (31.0)	41.4 (14.8)
3D	Initiator	10	51.7 (19.9)	20.7 (12.3)	29.7 (20.2)	51.7 (26.3)	58.7 (20.7)	32.2 (25.9)	40.8 (11.0)
2D	Discussant	10	52.0 (27.9)	31.0 (20.6)	26.0 (19.1)	45.0 (19.0)	42.0 (22.3)	37.5 (23.7)	38.9 (15.2)
3D	Discussant	10	50.5 (25.4)	31.0 (26.8)	28.0 (21.4)	48.5 (27.7)	41.0 (11.7)	33.0 (28.2)	38.6 (17.3)

[a] Note that the *Performance* sub-scale is labelled from 'Perfect' to 'Failure' i.e., a lower score is associated with better performance, and vice-versa.

Table 5
Mean and Standard deviation (within parenthesis) of Likert-scale measures common and unique to different participant *Roles*, grouped by Drawing Dimension and *Role*. Measures related to usefulness of the drawing tool (Usefulness DT) and virtual environment (Usefulness VE) during conversation were relevant to both *Role: Initiator* and *Role: Discussant* (prefixed with 'I.D.'). Measures only applicable to *Role: Initiator* were prefixed with 'I.' and include ratings on understanding of the video content (I.Understanding VC), perceptions of their explanation accuracy (I.Perception EA), and perceptions of the discussant's understanding (I.Perception DU). Measures only applicable to *Role: Discussant* were prefixed with 'D.' and include ratings on understanding the Initiators Explanation (D.Understanding IE), and Perceptions on their recounting accuracy of the initiators explanation (D.Perception RA).

Drawing dimension	Role	N	I.Understanding VC [0–5]	I.Perception EA [0–5]	I.Perception DU [0–5]	D.Understanding IE [0–5]	D.Perception RA [0–5]	I.D.Usefulness DT [0–5]	I.D.Usefulness VE [0–5]
2D	Initiator	10	3.5 (0.7)	3.6 (0.5)	3.0 (1.2)	–	–	3.7 (1.0)	3.3 (0.9)
3D	Initiator	10	3.8 (1.0)	3.7 (0.9)	3.8 (1.0)	–	–	4.4 (0.5)	3.6 (1.1)
2D	Discussant	10	–	–	–	3.8 (0.7)	3.6 (0.9)	4.0 (0.8)	3.3 (1.0)
3D	Discussant	10	–	–	–	4.2 (0.6)	4.0 (0.6)	4.4 (0.5)	3.9 (0.8)

(I.D.Usefulness VE). Finally, we found no evidence using CLMM that Drawing Dimension was a significant predictor of any of the measures unique to participant *Roles* — I.Understanding VC ($b = 0.816,\ z = 1.296,\ p = 0.195$), I.Perception EA ($b = 0.326,\ z = 0.616,\ p = 0.537$), I.Perception DU ($b = 1.180,\ z = 1.892,\ p = 0.058$), D.Understanding IE ($b = 0.700,\ z = 1.258,\ p = 0.209$), and D.Perception RA ($b = 0.668,\ z = 1.123,\ p = 0.262$).

7. Discussion & future work

Our study aimed to better understand the interconnection of actions and speech that interlocutors employ when supporting spatial dialogue with 2D or 3D drawings in VR. In this section, we unpack our findings to discuss the key differences and similarities between 2D and 3D drawings in supporting spatial dialogue, and highlight pertinent design considerations unique to each modality.

7.1. Interlocutors' actions, speech & their connections

The model created through the use of ENA on our coded data revealed distinct trends between the connections of interlocutor actions and speech when using 2D or 3D drawing during spatial dialogue in VR. The strongest (thickest edge) connection observed in both 2D and 3D conditions related to the use of drawings to illustrate verbal discussion points (*Drawing Mediated Speech*) and kinesic movements that referred to created drawings (*Deictic Drawing Reference*). While expected, the presence and strength of this connection reassure us that both 2D and 3D drawing served the intended purpose of supporting communication (Kang et al., 2015) by creating new spatial reference frames that were referred to through deictic expressions to support spatial dialogue.

Examining Fig. 2, depicting the connections of coded elements during 2D drawing use, suggests relatively strong connections between *Drawing Mediated Speech* and the kinesic movements related to *Deictic Tracing* (movements that trace paths along a created drawing) and *Pictograph* (movements illustrating a real object/referent). In addition, we also observe strong connections between *Drawing Mediated Speech* and drawing when referring to other created drawings (*Illustration Referencing Drawing)*. The connections between *Drawing Mediated Speech* and *Pictograph*, and between *Drawing Mediated Speech* and *Illustration Reference Drawing*, reveal the use of additional illustrations, through gestures (pictographs) and other drawings (other drawing referents), to reinforce communication during 2D drawing tool use. Similar connections were not observed in our 3D condition, as illustrated in Figs. 3 and 4.

Observations of our video recordings indicate that use of additional illustrations in the 2D condition was caused by two primary factors: the need to depict multiple orthographic illustrations and relate them to a single (or multiple) isometric view of the spatial concept (*Illustration Referencing Drawing*), and the limited drawing real-estate (whiteboard space) — exacerbated by the visual clutter caused by multiple orthographic illustrations — prompting users to make up for the lack of space by using the area outside the virtual whiteboard (*Pictograph*). Addressing the latter challenge of limited drawing space, has been explored in prior work, such as with solutions employing infinite virtual whiteboards enabling pan and zoom functions (Grønbæk et al., 2024). The viability of such a solution for spatial dialogue was reinforced by participant comments in S3:'… if you could, for example, pinch and expand the size of the whiteboard…'. However, navigating and interacting with an infinite 2D whiteboard requires additional controls/gestures that could increase workload and influence communication gestures employed when using 2D drawings for spatial dialogue in VR. Additionally, our findings (Fig. 2) suggest that minimizing illustrative gestures through solutions such as an infinite 2D canvas may heighten challenges in referring to related drawings due to the increased number of drawings afforded on the infinite 2D canvas, and the reduced spatial memory performance in 2D when compared to 3D representations (Tavanti and Lind, 2001). This relationship can be observed in the connections depicted in Fig. 2, where an infinite whiteboard would reduce connections between *Drawing Mediated Speech* and *Pictograph* by allowing space for more related drawings, but would exacerbate challenges with connections between *Drawing Mediated Speech* and *Illustration Referencing Drawing* by requiring interlocutors to find and refer to the increased number of related drawings. These findings highlight the need to consider the nuanced interconnection of actions observed during drawing supported spatial dialogue when designing solutions. An example solution to the limited drawing space issue, while minimizing impacts on spatial memory when using 2D drawing tools, is an infinite whiteboard with segmentation features that enable individual 2D illustrations to be segmented and organized in 3D space.

In the case of 3D drawing use for spatial dialogue, we observed stronger connections between *Drawing Mediated Speech*, deictic expressions referring to the virtual environment (*Deictic Environment Reference*), and drawings situated in relevant virtual environment locations (*Environment Situated Drawing*). These indicate a higher degree of awareness of the space and virtual environment around interlocutors when using 3D drawing tools. This awareness enables interlocutors to supplement and ground their explanations in relevant virtual surroundings when using 3D drawing tools. However, heightened awareness in 3D also increases the demand for creating virtual environments relevant to the dialogue topic/task to mitigate distractions (Bian et al., 2018). For example, a participant in S1 mentioned 'I don't know how effective they [the environment] were [for communicating], because I didn't, you know, create the content [environment] for the instructions. The need to prepare specific virtual environments could couple the drawing activity to the prepared environment, thereby undermining salient characteristics of drawing as a spontaneous and performative communicative practice (Snyder, 2013). Additionally, this finding also hints at possible challenges in employing 3D drawing for remote collaborations in mixed/augmented reality (MR/AR), where referencing or grounding illustrations in the *physical* surroundings would be impossible due to differences in the physical spaces of collaborators (Sra et al., 2018).

These findings suggest that 3D drawing tools can enable rich illustrations that leverage spatial awareness to support spatial dialogue in VR, and possibly in co-located MR/AR. However, open challenges remain in enabling collaborators to create meaningful shared environmental content and references, while maintaining the flexible and impromptu support that drawing offers for spatial dialogue.

In contrast, we observed only a weak connection between *Deictic Environment Reference* and *Drawing Mediated Speech*, and no connections between *Deictic Environment Reference* and other coded elements in our 2D condition. This suggests a lack of actions performed to communicate the association between the conceptual illustrations using 2D drawings, with the virtual counterparts of real-world referents in the virtual environment. Additionally, despite alleviating discrepancies between 2D and 3D modalities in environmentally situating drawings via a movable 2D whiteboard, we found no codes for *Environment Situated Drawing* in the 2D condition. We argue that this was because the virtual whiteboard becomes the focus of attention for the communication task (Syiem et al., 2021), and subsequently blinds individuals to the use of referents outside this *perceptual group* (Treisman, 1982; Syiem et al., 2024). Focusing on the whiteboard may also prove advantageous in reducing surrounding distractions when highly relevant virtual environments are unavailable. This finding also suggests that 2D drawing tools may be more suitable than 3D tools in remote MR/AR collaborations, as focus on an aligned whiteboard/2D-surface (Grønbæk et al., 2024) and drawings would prevent use of inconsistent reference frames across the different physical spaces of collaborators. In cases where the surrounding environment can support spatial dialogue, such as in *co-located* MR/AR or in relevant virtual environments, methods to minimize excessive focus on the whiteboard could benefit spatial dialogue supported by 2D drawings. For instance, we can take lessons from interlocutors in our 3D conditions that positioned themselves face-to-face while discussing/creating drawings between them. This arrangement enabled interlocutors to view the drawings, while being visually aware of each other and the virtual environment. Similar solutions can be explored for 2D drawing in VR; for instance, with a transparency-adjustable board, enabling interlocutors to decide how salient the virtual environment will be while anchoring focus on the virtual whiteboard.

Finally, we observed stronger connections between *Drawing Mediated Speech* and *Prompted Spatial Organization* in the 3D condition. Examination of our video recordings indicates that the connections related to *Prompted Spatial Organization* not only suggest explicit use of 3D space when communicating about 3D illustrations, but also spotlight issues related to *aligning different interlocutors' 3D perspective views* when using drawings (*Drawing Mediated Speech*) or deictic gestures (Deictic Drawing Reference) for spatial dialogues (Pouliquen-Lardy et al., 2016). While 3D drawings enable complex representations that can contain information typically spread across multiple orthographic viewpoints in 2D (Arora et al., 2023) — this complexity introduces challenges in establishing a common frame of reference and viewpoint for discussing 3D drawings. This challenge was observed despite high self-reports by participants on their experience with 3D thinking & reasoning (Barrera Machuca et al., 2019) (M: 4.2, SD: 0.6, out of 5), and suggests the need for solutions to enable effective means of referring to specific elements of a complex 3D illustration, while simultaneously communicating the respective viewpoint, regardless of user expertise. Deictic gestures and visible avatar head positions provide a means to establish a shared frame of reference when discussing 3D drawings, but are not sufficient to visualize the *exact* orthographic view that an interlocutor may be referring to within a 3D drawing — a problem that is not present when communicating using 2D drawings.

7.2. Opinions, performance, & workload

Despite the differences found in interlocutors' actions and speech during communication using 2D and 3D drawings, we found no significant differences in completion time and the total time spent drawing during spatial dialogue between the two conditions. Interlocutors spent similar amounts of time conversing in the 2D condition (M: 681.5s, SD: 184.1s) and in the 3D condition (M: 643.3s, SD: 189.8s). Additionally, interlocutors spent similar amounts of time drawing illustrations in 2D (M: 45.6.5s, SD: 30.1s) and 3D (M: 44.7s, SD: 29.4s) conditions. This is surprising, as we expected users to spend more time creating 3D drawings and verifying their accuracy from multiple 3D viewpoints given its novelty.

A possible explanation is that interlocutors could always fall back on creating familiar 2D drawings in 3D space when necessary. For instance, participants in S8 mentioned 'even though we are doing this [in] 3D, I would try to express those [concepts] in a 2D world, and in S3 'I was thinking that you could still do the 2D on the 3D right?'. These remarks were supported by the number of 2D communicative illustrations (97/182 instances or 53%) made during the 3D condition, i.e., the number of instances when interlocutors explained a task concept using a 2D perspective (single viewpoint) of a drawing made in the 3D condition. We count communicative illustrations, in place of individual drawings, due to the iterative, or developing/ordered (Tversky, 2002; Pache et al., 2001; Snyder, 2013), nature in which drawings were constructed during communication in our study[12]. These findings, along with our ENA analysis (Section 6.1), suggest that interlocutors did not solely depend on familiar 2D drawing actions when using the 3D drawing tool but often leveraged the available 3D capabilities. While the ability to draw 2D illustrations using 3D drawing tools may seemingly place 3D at an advantage for supporting spatial dialogue, participant opinions were largely mixed during our interviews. For instance, participants in S4 expressed that they found it hard to take advantage of 3D drawing as 2D was 'more like what I'm used to traditionally, while participants in S3 mentioned that they preferred 3D, because using 2D was limiting — 'like bringing that [2D] real-world limitation into the VR world seems pointless to me. Other dyads argued for having the option to use both 2D and 3D drawing; (S1 participant) 'because you want to explain things in 2D as well as 3D, right?'.

One benefit of using 2D over 3D drawings in VR lies in the added shared reference frame of the virtual whiteboard. Our analysis of the video recordings and ENA models suggests that the whiteboard not only served as a canvas but as a common anchor that interlocutors could spatially organize around in a familiar manner. The lack of familiar anchors during 3D drawing tools use, meant that interlocutors performed more movement (both prompted and unprompted) to meaningfully situate and orient themselves. This is indicated in Fig. 4 which shows more connections between both *Prompted Spatial Organization* and *Unprompted Spatial Organization* to *Drawing Mediated Speech* for 3D when compared to 2D. Challenges related to spatial organization in 3D were also highlighted in our interviews. For example a participant in S4 elaborates on the challenges of spatial organization in 3D — 'I guess trying to observe, like, their instructions when you're like, facing them, everything they do is backwards. So, you can just sort of like, rotate around, and then you're like, facing it with them, but then you're not, like, looking at them while they're trying to instruct…'. However, the same freedom to spatially organize in 3D was seen as beneficial by other dyads. S7 on the use of 3D drawings for communication — 'it is more interactive … it would be more fun, being able to move around and see what you are doing than just at a, uh, a screen'.

The mixed opinions expressed during our interviews echo the similar responses we received regarding how useful participants found the different drawing modalities. Participants rated the usefulness of 2D drawings to support communication marginally lower (M: 3.85, SD: 0.93) than 3D drawings (M: 4.4, SD: 0.50) — which was a statistically non-significant difference. Participants also reported similar usefulness of the VR environment in the 2D (M: 3.3, SD: 0.97) and 3D (M: 3.75,

[12] An example of such behaviour can be seen in the video provided in our supplementary material.

SD: 1.02) conditions. However, these responses were not indicative of the lack of actions performed during the 2D condition to refer to, or leverage, the virtual environment (see Fig. 2).

Finally, no significant differences were found in subjective workload and post-test scores between the 2D and 3D conditions in our experiment. While our experiment lacks an adequate sample size to accurately assess statistical differences, our quantitative data present preliminary insights and provide additional context to the findings related to our qualitative data. The lack of significant results does not imply an absence of an effect of Drawing Dimension on task load and post-test scores. However, the similarity in measures hints at the *potential* for adopting the different communication strategies afforded by 2D and 3D drawing without excessively increasing subjective workload or negatively impacting content understanding. Future work is needed to investigate the extent to which the different actions and speech employed during 2D and 3D drawing use for spatial dialogue impact workload, content understanding, and retention.

8. Limitations & future work

While our study demonstrates the effects of complementing spatial dialogue with different drawing modalities (2D or 3D) in VR on interlocutors' actions and speech, there are limitations that should be considered when interpreting our findings. For instance, our study employs a task related to structural geology that involves abstract 3D concepts and their relationship with environmental contexts (evinced by the importance of fieldwork in geoscience (Gallagher et al., 2021)). While communicating such 3D concepts is widespread in fields such as chemistry and physics, the insights from our study may not apply to spatial dialogue that concerns more concrete 3D structures, such as in applications for product design or urban planning.

An additional limitation arises from our use of controller-based drawing and navigation in our collaborative VR system. Controller-based interactions were used instead of hand-based interactions to reduce non-communicative hand gestures that could influence communication behaviour during the experimental task (see Section 3). This was crucial as our study focuses primarily on interlocutor actions and speech during spatial dialogue. However, the use of controllers limits the articulation of avatar hands during communication, and future work is needed to explore the trade-offs between controller and hand-based interactions during drawing supported spatial dialogue in VR. We also considered pen-like stylus controllers for VR (such as the MX Ink) as they provide more familiar drawing interactions when compared to VR controllers. However, VR stylus controllers were not available to us during this study, and are generally less accessible than VR controllers — requiring independent procurement, unlike VR controllers or inbuilt hand-recognition that are included with most modern VR headsets. VR stylus controllers may also introduce challenges for navigating the VR environment, requiring unconventional control mappings for teleporting and translating the user's position. Future work is needed to investigate the effects of different controller types on interaction affordances during spatial dialogue in VR.

Additionally, to control for the effect of avatar appearance on communication (Aseeri and Interrante, 2021), our VR collaboration system only enables two different avatars (see Section 3). However, such an implementation may not be ecologically valid as modern VR applications enable avatar customization. Further, our avatars movements are solely based on three trackers, namely the headset and the two VR controllers. Use of additional trackers may improve their realism and representations, but comes at the cost of more complicated or invasive technology, such as full-body tracking suits or computer vision based body recognition integration with VR. As such, future work is needed to investigate the influence of, and best practices for, avatar representations for spatial dialogue tasks in VR.

Lastly, a limitation in our study arose from the need to balance interaction complexity with the distinct affordances of 2D and 3D drawing spaces. While 3D drawing tools in VR afford natural movements to draw anywhere in infinite space, enabling equivalent 2D infinite canvases in VR would require additional control/gesture mappings for navigating the 2D canvas. Such differences in interaction complexities between the two conditions could influence task-load and communication strategies employed. Alternatively, we considered limiting the 3D drawing space by bounding it to a particular sub-section of space to control for the discrepancies between 2D and 3D drawing space availability. However, this would remove a core affordance of 3D drawings (i.e., spatial freedom (Arora et al., 2023)), and result in a contrived scenario that does not reflect how interlocutors typically behave and communicate when using 3D drawings for spatial dialogue in VR. As a compromise, we chose to provide a large enough 2D whiteboard to comfortably accommodate all task-related illustrations, enabling us to maintain similar interaction complexity while preserving core affordances of the different modalities. However, this leaves open questions regarding the use of infinite 2D canvases for communicating spatial concepts in VR.

9. Conclusion

In this paper, we investigated the effects of 2D and 3D drawing modalities on communication strategies (actions and speech) employed by interlocutors during spatial dialogue in VR. Our analysis revealed that participants showed similar task completion times and reported comparable levels of workload when using 2D or 3D drawing tools to support spatial dialogue in VR. However, participants expressed mixed preferences during our interviews, highlighting benefits and challenges of both 2D and 3D drawing modalities for communication. These findings suggest that both 2D and 3D drawings tools provide viable means of supporting spatial dialogue in VR, but underscore the need for flexible design considerations around drawing modalities in VR to accommodate a wide range of communication needs and preferences.

Despite similarities in performance, we found significant differences in the ways interlocutors act and verbally communicate between the two modalities. 2D drawings were found to facilitate more traditional approaches for conveying spatial concepts by structuring communication around a familiar 2D drawing surface. However, the need to produce multiple orthographic views to represent 3D concepts led to a visually cluttered canvas and prompted interlocutors to rely on gestures to supplement their communication. In contrast, 3D drawings allowed participants to create more complex illustrations, taking advantage of the surrounding virtual space. While this provided greater flexibility and expressiveness, it also introduced challenges in aligning the different spatial perspectives of interlocutors. These findings illuminate the various communication strategies employed by interlocutors when using the different drawing modalities, and provide rich insights into the challenges, benefits and affordances of the drawing modalities in supporting spatial dialogue.

Our findings contribute to a better understanding of how 2D and 3D drawing modalities influence communication strategies in VR. We show that both drawing modalities offer distinct advantages and prompt unique actions and speech for supporting spatial dialogue in collaborative VR. However, future VR applications must strike a balance between structure and flexibility, offering tools that address contemporary needs for communicating complex 3D spatial structures, while maintaining the intuitive use of gestures, speech, and space seen in face-to-face communication.

CRediT authorship contribution statement

Brandon Victor Syiem: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Selen Türkay:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Cael Gallagher:** Writing – review & editing, Conceptualization. **Christoph Schrank:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Selen Türkay reports that financial support was provided by Australian Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was supported by an Australian Research Council Discovery Project Grant (ARC DP210103889) funded by the Australian Government.

Appendix A

A.1. Custom Likert-scale surveys: perceived communication effectiveness

All survey items were on a 5-point Likert-scale with responses (in ascending order): 'Strongly Disagree', 'Disagree', 'Neither Agree nor Disagree', 'Agree', and 'Strongly Agree'. We provide the survey items for the Initiator and Discussant separately. Text presented in bold was also bold in the survey presented to the participants.

Initiator.

- I understood the **content presented in the video**.
- I feel that I explained **my understanding** of the video content accurately to my collaborator.
- I feel that my collaborator understood **my explanation** accurately.
- The **drawing tools** in virtual reality helped me with my explanation.
- The **virtual environment** helped me with my explanation.

Discussant.

- I feel that I understood my **collaborator's explanation** accurately.
- I feel that I accurately **recounted** my collaborator's explanation.
- The **drawing tools** in virtual reality helped me with my explanation.
- The **virtual environment** helped me with my explanation.

A.2. Interview questions

Interview questions were used to gather insights into participant opinions on the overall experience of communicating spatial concepts in VR using 2D or 3D drawing tools.

- How was your overall experience using the VR application to communicate? Did you encounter any difficulties? (follow-up: please elaborate)
- Did you find any differences when trying to communicate using 2D or 3D drawing tools? Did you have a preference for 2D or 3D drawing tools when creating drawings, and when viewing drawings, for communication? (follow-up: elaborate on your preferences)
- Did the virtual environment affect your communication or creation of illustrations? (follow-up: please elaborate on how)

A.3. Example screenshots from our task videos

As detailed in Section 4.4, one participant in the dyad for each experimental condition was asked to watch a video presenting structural geology related spatial concepts and communicate these concepts to the other participant using our VR system. To illustrate the spatial nature of the task we used, we provide a few example screenshots from the task videos in Fig. A.5.

Note that we do not have permission to share the entirety of these lecture videos. To further clarify, our participants were **not** asked to

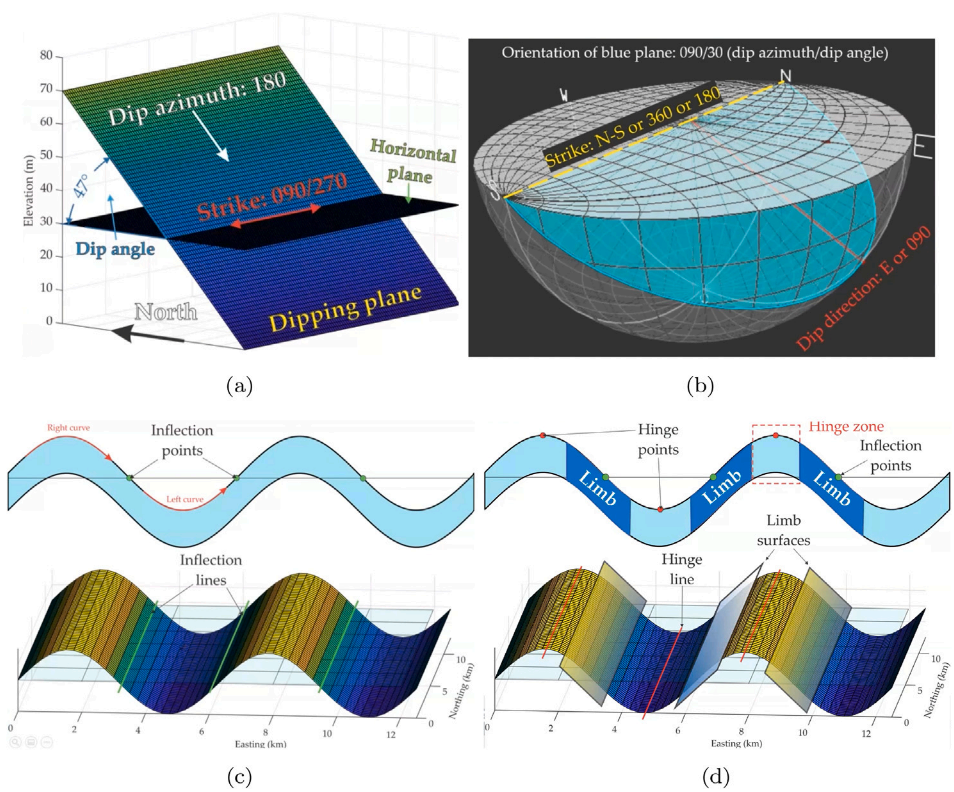


Fig. A.5. Screenshots from our Task Videos. Screenshots (a) and (b) were taken from our task video related to 'Dip & Strike'. The bottom images, (c) and (d), were part of the video lecture related to the geological structure of 'Folds'. Note that the 3D hemispherical projection presented in (b) was produced with Stereonet V. 11 (Allmendinger et al., 2011; Cardozo and Allmendinger, 2013).

Table A.6
Post-test questions and scoring criteria used to assess understanding of the prescribed topic of 'Dip & strike'. Each question is followed by a sub-table; with each column depicting the score and the criteria for receiving that score.

Dip & Strike				
1: What is strike? How do you report strike in relation to the cardinal directions (North, East, South & West)?				
0	1	2	3	4
Mentions nothing relevant	Mentions a plane **or** mentions a line	Mentions a plane **and** mentions a line	Describes the strike as a horizontal line on a plane *or* As a line on a plane **and** correctly describes how it is reported (i.e, two opposing cardinals)	Describes that the strike is a horizontal line on a plane **and** describes correctly how the cardinal directions are reported (i.e, two opposing cardinals)
2: What is dip direction? How is dip direction reported numerically?				
0	1	2	3	4
Mentions nothing relevant	Mentions that dip direction is a line that goes 'down' a plane	Mentions that dip direction is a line that goes 'down' a plane **and** correctly describes how to report it (i.e, three digits from 000 to 360)	Describes dip direction in relation to strike (as being perpendicular to it) **or** as the direction that water trickles down a surface	Describes dip direction as per 3 **and** also correctly describes how to report it.
3: Draw an illustration of strike and dip direction (dip azimuth and dip angle) with labels.				
0	1	2	3	4
Draws nothing relevant	Almost correctly draws and labels one	Correctly draws and labels one	Correctly draws and labels two	Correctly draws and labels three

Table A.7
Post-test questions and scoring criteria used to assess understanding of the prescribed topic of 'Geological Folds'. Each question is followed by a sub-table; with each column depicting the score and the criteria for receiving that score.

Geological Folds				
1: What is a fold hinge and fold limb?				
0	1	2	3	4
Mentions nothing relevant	Mentions a waveform	Mentions the maximum curvature **or** the point of inflections	Correctly describes the hinge as the point of maximum curvature **or** the limb as the straight part where you find the point of inflection	Correctly describes both
2: What are inflection points and lines? How are they related to fold limbs?				
0	1	2	3	4
Mentions nothing relevant	Almost correctly describes what a point of inflection is	Almost correctly describes what a point of inflection is **and** relates it to the limb	Correctly describes a point of inflection	Correctly describes a point of inflection **and** relates it to the limb of a fold.
3: Draw an illustration of fold hinge, limbs and inflection points and lines with labels.				
0	1	2	3	4
Draws nothing relevant	Almost correctly draws and labels one	Correctly draws and labels one	Correctly draws and labels two	Correctly draws and labels three

replicate the images that they observed in the task videos. Instead, participants were solely tasked **with explaining the concepts presented in the videos**.

A.4. Post-test questionnaires & scoring criteria: content understanding

While our study did not focus on content understanding, we employed post-test questionnaires to gain preliminary insights into content understanding. The topics addressed in our post-test questionnaire were all covered in the video prescribed to the initiator during our experiment. The specific structural geology topics used were; 1) Dip & Strike, and 2) Geological Folds. Questions and scoring criteria used in our study were developed in consultation with a senior academic in structural geology, and are presented in Tables A.6 and A.7.

Table A.8
As detailed in Section 5.1.1, ENA creates a normalized vector from the frequency of co-occurrence between each pair of elements/codes during a conversation. The average of these normalized values in each experimental condition represents the relative connection strength between different codes in that condition. This table presents the average normalized values for both 2D and 3D conditions in our experiment as determined by ENA. Values below the main diagonal represent the average connection strength between codes for the 2D condition. Values above the main diagonal represent the average connection strength between codes for the 3D condition.

3D / 2D	Drawing Mediated Speech	Illustration Referencing Drawing	Environment Situated Drawing	Prompted Spatial Organization	Unprompted Spatial Organization	Speaker Pose Change	Baton	Ideograph	Kinetograph	Pictograph	Spatial Gesture	Deictic Drawing Reference	Deictic Environment Reference	Deictic Tracing
Drawing Mediated Speech		0.014	0.119	0.068	0.189	0.265	0.039	0.058	0.040	0.071	0.107	0.508	0.119	0.243
Illustration Referencing Drawing	0.138		0	0.001	0	0.002	0.001	0	0	0.003	0	0.012	0	0.005
Environment Situated Drawing	0	0		0.011	0.033	0.029	0.002	0.015	0.007	0.011	0.026	0.049	0.043	0.036
Prompted Spatial Organization	0	0	0		0.019	0.024	0.002	0.007	0.002	0.009	0.013	0.040	0.011	0.018
Unprompted Spatial Organization	0.144	0.018	0	0		0.056	0.003	0.011	0.005	0.010	0.025	0.076	0.028	0.055
Speaker Pose Change	0	0	0	0	0		0.010	0.024	0.013	0.023	0.050	0.122	0.034	0.051
Baton	0.068	0	0	0	0.025	0		0.001	0	0.004	0.007	0.030	0.001	0.007
Ideograph	0.062	0.002	0	0	0.013	0	0.014		0.007	0.007	0.015	0.027	0.019	0.011
Kinetograph	0.019	0.007	0	0	0.004	0	0.003	0.001		0.003	0.004	0.011	0.009	0.007
Pictograph	0.198	0.005	0	0	0.021	0	0.016	0.016	0		0.017	0.061	0.013	0.022
Spatial Gesture	0.114	0.013	0	0	0.008	0	0.002	0.009	0.001	0.020		0.089	0.031	0.030
Deictic Drawing Reference	0.582	0.028	0	0	0.075	0	0.057	0.062	0.005	0.105	0.079		0.060	0.144
Deictic Environment Reference	0.046	0.002	0	0	0.003	0	0	0	0	0.008	0.009	0.015		0.036
Deictic Tracing	0.277	0.021	0	0	0.066	0	0.047	0.036	0.007	0.045	0.031	0.198	0.006	

Appendix B. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.ijhcs.2025.103725.

Data availability

The authors do not have permission to share data.

References

Allmendinger, R.W., Cardozo, N., Fisher, D.M., 2011. Structural Geology Algorithms: Vectors and Tensors. Cambridge University Press.

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., Shaffer, D., 2015. Look together: analyzing gaze coordination with epistemic network analysis. Front. Psychol. 6, 1016.

Arastoopour Irgens, G., Shaffer, D.W., Swiecki, Z., Ruis, A.R., Chesler, N.C., 2015. Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. Int. J. Eng. Educ..

Arora, R., Machuca, M.D.B., Wacker, P., Keefe, D., Israel, J.H., 2023. Introduction to 3d sketching. In: Interactive Sketch-Based Interfaces and Modelling for Design. River Publishers, pp. 151–177.

Aseeri, S., Interrante, V., 2021. The influence of avatar representation on interpersonal communication in virtual social environments. IEEE Trans. Vis. Comput. Graph. 27, 2608–2617.

Barrera Machuca, M.D., Stuerzlinger, W., Asente, P., 2019. The effect of spatial ability on immersive 3d drawing. In: Proceedings of the 2019 Conference on Creativity and Cognition, pp. 173–186.

Bian, Y., Yang, C., Zhou, C., Liu, J., Gai, W., Meng, X., Tian, F., Shen, C., 2018. Exploring the weak association between flow experience and performance in virtual environments. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12.

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., Shaffer, D.W., 2021. The mathematical foundations of epistemic network analysis. In: Advances in Quantitative Ethnography: Second International Conference, ICQE 2020. Springer, pp. 91–105. Malibu, CA, USA, February 1–3 Proceedings 2.

Bressler, D.M., Bodzin, A.M., Eagan, B., Tabatabai, S., 2019. Using epistemic network analysis to examine discourse and scientific practice during a collaborative game. J. Sci. Educ. Technol. 28, 553–566.

Cardozo, N., Allmendinger, R.W., 2013. Spherical projections with osxstereonet. Comput. Geosci. 51, 193–205. https://doi.org/10.1016/j.cageo.2012.07.021. https://www.sciencedirect.com/science/article/pii/S0098300412002646..

Christensen, R.H.B., 2018. Cumulative link models for ordinal regression with the r package ordinal. J. Stat. Softw. 35.

Clark, H.H., 1991. Grounding in communication. In: Perspectives on Socially Shared Cognition. American Psychological Association.

Coelho, H., Monteiro, P., Gonçalves, G., Melo, M., Bessa, M., 2022. Authoring tools for virtual reality experiences: a systematic review. Multimed. Tools Appl. 81, 28037–28060.

Coventry, K.R., Tenbrink, T., Bateman, J., 2009. Spatial Language and Dialogue, vol. 3. OUP Oxford.

D'Angelo, A.-L.D., Ruis, A.R., Collier, W., Shaffer, D.W., Pugh, C.M., 2020. Evaluating how residents talk and what it means for surgical performance in the simulation lab. Am. J. Surg. 220, 37–43.

de Belen, R.A.J., Nguyen, H., Filonik, D., Del Favero, D., Bednarz, T., 2019. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. AIMS Electron. Electr. Eng. 3, 181–223.

Dudley, J.J., Schuff, H., Kristensson, P.O., 2018. Bare-handed 3d drawing in augmented reality. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 241–252.

Duronto, P.M., Nishida, T., Nakayama, S.-I., 2005. Uncertainty, anxiety, and avoidance in communication with strangers. Int. J. Intercult. Relat. 29, 549–560.

Dwivedi, V., Bhatnagar, M., Tursunov, J., Rozinaj, G., 2024. Immersive virtual conference room: bridging the gap between remote collaboration and physical presence. In: 2024

31st International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, pp. 1–5.
Efron, D., 1941. Gesture and environment.
Ekman, P., Friesen, W.V., 1969. The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica 1, 49–98.
Elmoazen, R., Saqr, M., Tedre, M., Hirsto, L., 2022. A systematic literature review of empirical research on epistemic network analysis in education. IEEE Access 10, 17330–17348.
Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., Billinghurst, M., 2019. Revisiting collaboration through mixed reality: the evolution of groupware. Int. J. Hum.-Comput. Stud. 131, 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011
Gallagher, C., Turkay, S., Brown, R.A., 2021. Towards designing immersive geovisualisations: literature review and recommendations for future research. In: Proceedings of the 33rd Australian Conference on Human-Computer Interaction, pp. 307–326.
Gasques, D., Johnson, J.G., Sharkey, T., Feng, Y., Wang, R., Xu, Z.R., Zavala, E., Zhang, Y., Xie, W., Zhang, X., et al., 2021. Artemis: a collaborative mixed-reality system for immersive surgical telementoring. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14.
Gifford, J., 2022. Remote working: unprecedented increase and a developing research agenda. Hum. Resour. Dev. Int. 25, 105–113.
Grønbæk, J.E.S., Sánchez Esquivel, J., Leiva, G., Velloso, E., Gellersen, H., Pfeuffer, K., 2024. Blended whiteboard: physicality and reconfigurability in remote mixed reality collaboration. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–16.
Hayward, W.G., Tarr, M.J., 1995. Spatial language and spatial representation. Cognition 55, 39–84.
He, Z., Du, R., Perlin, K., 2020. Collabovr: a reconfigurable framework for creative collaboration in virtual reality. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, pp. 542–554.
Irlitti, A., Latifoglu, M., Hoang, T., Syiem, B.V., Vetere, F., 2024. Volumetric hybrid workspaces: interactions with objects in remote and co-located telepresence. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–16.
Israel, J.H., Wiese, E., Mateescu, M., Zöllner, C., Stark, R., 2009. Investigating three-dimensional sketching for early conceptual design—results from expert discussions and user studies. Comput. Graph. 33, 462–473.
Jacob, R.J.K., 1995. Eye tracking in advanced interface design. Virtual environments and advanced interface design 258, 2.
Jang, S.-A., Wakefield, G., Lee, S.-H., 2017. Incorporating kinesthetic creativity and gestural play into immersive modeling. In: Proceedings of the 4th International Conference on Movement Computing, pp. 1–8.
Kang, S., Tversky, B., Black, J.B., 2015. Coordinating gesture, word, and diagram: explanations for experts and novices. Spat. Cogn. Comput. 15, 1–26.
Kingsley, L.J., Brunet, V., Lelais, G., McCloskey, S., Milliken, K., Leija, E., Fuhs, S.R., Wang, K., Zhou, E., Spraggon, G., 2019. Development of a virtual reality platform for effective communication of structural data in drug discovery. J. Mol. Graph. Model. 89, 234–241.
Li, J., Vinayagamoorthy, V., Williamson, J., Shamma, D.A., Cesar, P., 2021. Social VR: a new medium for remote communication and collaboration. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–6.
Liddell, T.M., Kruschke, J.K., 2018. Analyzing ordinal data with metric models: what could possibly go wrong? J. Exp. Soc. Psychol. 79, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009. https://www.sciencedirect.com/science/article/pii/S0022103117307746..
Marquart, C., Hinojosa, C., Swiecki, Z., Shaffer, D.W., 2018. Epistemic network analysis. http://app.epistemicnetwork.org/login.html..
McNeill, D., 1992. Hand and mind1. Adv. Vis. Semiot. 351.
McVeigh-Schultz, J., Kolesnichenko, A., Isbister, K., 2019. Shaping pro-social interaction in VR: an emerging design framework. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12.
Nash, P., Shaffer, D.W., 2013. Epistemic trajectories: mentoring in a game design practicum. Instr. Sci. 41, 745–771.
Olaosebikan, M., Aranda Barrios, C., Kolawole, B., Cowen, L., Shaer, O., 2022. Identifying cognitive and creative support needs for remote scientific collaboration using VR: practices, affordances, and design implications. In: Creativity and Cognition, pp. 97–110.
Osborne, A., Fielder, S., Mcveigh-Schultz, J., Lang, T., Kreminski, M., Butler, G., Li, J.V., Sanchez, D.R., Isbister, K., 2023. Being social in VR meetings: a landscape analysis of current tools. In: Proceedings of the 2023 ACM Designing Interactive Systems Conference, pp. 1789–1809.
Oti, A., Crilly, N., 2021. Immersive 3d sketching tools: implications for visual thinking and communication. Comput. & Graph. 94, 111–123.
Ouverson, K.M., Gilbert, S.B., 2021. A composite framework of co-located asymmetric virtual reality. Proc. ACM Hum.-Comput. Interact. 5, 1–20.
Pache, M., Römer, A., Lindemann, U., Hacker, W., 2001. Sketching behaviour and creativity in conceptual engineering design. In: International Conference on Engineering Design ICED01, Glasgow, August, pp. 21–23.
Pache, M.W., 2005. Sketching for Conceptual Design: Empirical Results and Future Tools, Ph.D. thesis, Citeseer.
Piumsomboon, T., Lee, Y., Lee, G., Billinghurst, M., 2017. Covar: a collaborative virtual and augmented reality system for remote collaboration. In: SIGGRAPH Asia 2017 Emerging Technologies, pp. 1–2.
Pollick, A.S., De Waal, F.B.M., 2007. Ape gestures and language evolution. Proc. Natl. Acad. Sci. 104, 8184–8189.
Porter, S.R., Whitcomb, M.E., Weitzer, W.H., 2004. Multiple surveys of students and survey fatigue. New Dir. Inst. Res. 2004, 63–73.
Pouliquen-Lardy, L., Milleville-Pennel, I., Guillaume, F., Mars, F., 2016. Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. Virtual Reality 20, 213–220.
Prieur, J., Barbu, S., Blois-Heulin, C., Lemasson, A., 2020. The origins of gestures and language: history, current advances and proposed theories. Biol. Rev. 95, 531–554, .
Rolim, V., Ferreira, R., Lins, R.D., Găsević, D., 2019. A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. Int. High. Educ. 42, 53–65.
Rovida, E., 2012. Machines and Signs: A History of the Drawing of Machines, vol. 17. Springer Science & Business Media.
Schäfer, A., Reis, G., Stricker, D., 2022. A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems. ACM Comput. Surv. 55, 1–27.
Schueler, A., Wesslein, A.-K., 2022. Investigating the effect of deictic tracing on multimedia learning. Learn. Instr. 77, 101525.
Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A.A., Mislevy, R., 2009. Epistemic network analysis: a prototype for 21st-century assessment of learning. Int. J. Learn. Media 1, 33–53.
Shaffer, D.W., Collier, W., Ruis, A.R., 2016. A tutorial on epistemic network analysis: analyzing the structure of connections in cognitive, social, and interaction data. J. Learn. Anal. 3, 9–45.
Smith, H.J., Neff, M., 2018. Communication behavior in embodied virtual reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12.
Snyder, J., 2011. Image-enabled discourse: a conceptual framework. In: Intl. Visual Soc. Association (IVSA) Conference.
Snyder, J., 2013. Drawing practices in image-enabled collaboration. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 741–752.
Snyder, J., 2014. Visual representation of information as communicative practice. J. Assoc. Inf. Sci. Technol. 65, 2233–2247.
Sra, M., Mottelson, A., Maes, P., 2018. Your place and mine: designing a shared VR experience for remotely located users. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 85–97.
Suwa, M., Tversky, B., 1996. What architects see in their sketches: implications for design tools. In: Conference Companion on Human Factors in Computing Systems, pp. 191–192.
Syiem, B.V., Türkay, S., 2024. A systematic exploration of collaborative immersive systems for sense-making in STEM. Behav. Inf. Technol. 1–30.
Syiem, B.V., Kelly, R.M., Goncalves, J., Velloso, E., Dingler, T., 2021. Impact of task on attentional tunneling in handheld augmented reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14.
Syiem, B.V., Kelly, R.M., Dingler, T., Goncalves, J., Velloso, E., 2024. Addressing attentional issues in augmented reality with adaptive agents: possibilities and challenges. Int. J. Hum.-Comput. Stud. 103324.
Tan, Y., Swiecki, Z., Ruis, A.R., Shaffer, D., 2024. Epistemic Network Analysis and Ordered Network Analysis in Learning Analytics. Springer Nature Switzerland, Cham, pp. 569–636. https://doi.org/10.1007/978-3-031-54464-4_18
Tavanti, M., Lind, M., 2001. 2d VS 3d, implications on spatial memory. In: IEEE Symposium on Information Visualization. INFOVIS 2001. IEEE, pp. 139–145.
Tenbrink, T., Andonova, E., Schole, G., Coventry, K.R., 2017. Communicative success in spatial dialogue: the impact of functional features and dialogue strategies. Lang. Speech 60, 318–329.
Thompson, M., Bilgin, C.U., Eberhart, R., Cho, L., Anteneh, M., Klopfer, E., 2021. Visualizing the collaborative problem solving process in an immersive cross platform game. In: 2021 7th International Conference of the Immersive Learning Research Network (iLRN). IEEE, pp. 1–6.
Treisman, A., 1982. Perceptual grouping and attention in visual search for features and for objects. J. Exp. Psychol. Hum. Percept. Perform. 8, 194.
Tversky, B., Lozano, S.C., Heiser, J., Lee, P.U., Daniel, M.-P., 2005. Explanations in diagram, word, and gesture. In: AAAI Spring Symposium: Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance, pp. 105–107.
Tversky, B., 2002. What do sketches say about thinking. In: 2002 AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University. AAAI Technical Report SS-02-08, vol. 148. pp. 151.
Uz-Bilgin, C., Thompson, M., Anteneh, M., 2020. Exploring how role and background influence through analysis of spatial dialogue in collaborative problem-solving games. J. Sci. Educ. Technol. 29, 813–826.
Ververidis, D., Nikolopoulos, S., Kompatsiaris, I., 2022. A review of collaborative virtual reality systems for the architecture, engineering, and construction industry. Architecture 2, 476–496.
Wang, R., Olson, B., Vaidyanathan, P., Bailey, R.J., Alm, C.O., 2019. Fusing dialogue and gaze from discussions of 2d and 3d scenes. In: Adjunct of the 2019 International Conference on Multimodal Interaction, https://doi.org/10.1145/3351529.3360661
Wei, X., Jin, X., Fan, M., 2022. Communication in immersive social virtual reality: a systematic review of 10 years' studies. In: Proceedings of the Tenth International Symposium of Chinese CHI, pp. 27–37.
Wu, H., Wang, J., 2016. A visual attention-based method to address the midas touch problem existing in gesture-based interaction. Vis. Comput. 32, 123–136.
Zhang, L., Agrawal, A., Oney, S., Guo, A., 2023. VRGIT: a version control system for collaborative content creation in virtual reality. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–14.