"Do I Just Tap My Headset?": How Novice Users Discover Gestural Interactions with Consumer Augmented Reality Applications

ANONYMOUS AUTHOR(S)

A variety of consumer Augmented Reality (AR) applications have been released on mobile devices and novel immersive headsets over the last five years, creating a breadth of new AR-enabled experiences. However, these applications, particularly those designed for immersive headsets, require users to employ unfamiliar gestural input and adopt novel interaction paradigms. To better understand how everyday users discover gestures and classify the types of interaction challenges they face, we observed how 25 novices from diverse backgrounds and technical knowledge used four different AR applications requiring a range of interaction techniques. A detailed analysis of gesture interaction traces showed that users struggled to discover the correct gestures, with the majority of errors occurring when participants could not determine the correct sequence of actions to perform or could not evaluate their actions. To further reflect on the prevalence of our findings, we carried out an expert validation study with 8 professional AR designers, engineers, and researchers. We discuss implications for designing discoverable gestural input techniques that align with users' mental models, inventing AR-specific onboarding and help systems, and enhancing system-level machine recognition.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: augmented reality; mixed reality; gesture discoverability; help-seeking

ACM Reference Format:
Anonymous Author(s). 2023. *"Do I Just Tap My Headset?"*: How Novice Users Discover Gestural Interactions with Consumer Augmented Reality Applications. In . ACM, New York, NY, USA, 30 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The exploration and implementation of Augmented Reality (AR) applications have witnessed significant growth [5] across various domains, including gaming (e.g., Pokemon Go [53]), digital shopping [17], and consumer-grade, task-based AR applications (e.g., *Insight Lungs* [4], *Google Maps* [65], *Holomeeting* [2], etc.). Despite this growth, the widespread adoption of AR applications remains a challenge due to the limited familiarity of many consumers with this cutting-edge technology [11, 69]. Unlike conventional interfaces such as WIMP (Windows, Icons, Menus, Pointer) and interactions based on keyboards/mice, AR introduces users to a myriad of novel interaction paradigms, including geolocated egocentric navigation and mid-air gestures, where users can interact with the user interface without touching or holding a physical device [16]. Users must transfer their understanding of desktop and mobile computing to seamlessly interact with interfaces that overlay onto the real world. The complexity intensifies in the domain of immersive AR hosted on head-mounted displays (e.g., Microsoft Hololens, Magic Leap), where sensors and cameras are used to seamlessly integrate 3D virtual objects into the user's immediate physical surroundings, blurring the boundaries between the digital and physical realms [8, 17, 42].

A growing body of HCI research is proposing novel gestural interaction techniques for a variety of AR environments[20, 24, 63, 72, 95, 112, 117]. Focus has mostly been on improving the accuracy and efficiency of gesture recognizers

Manuscript submitted to ACM

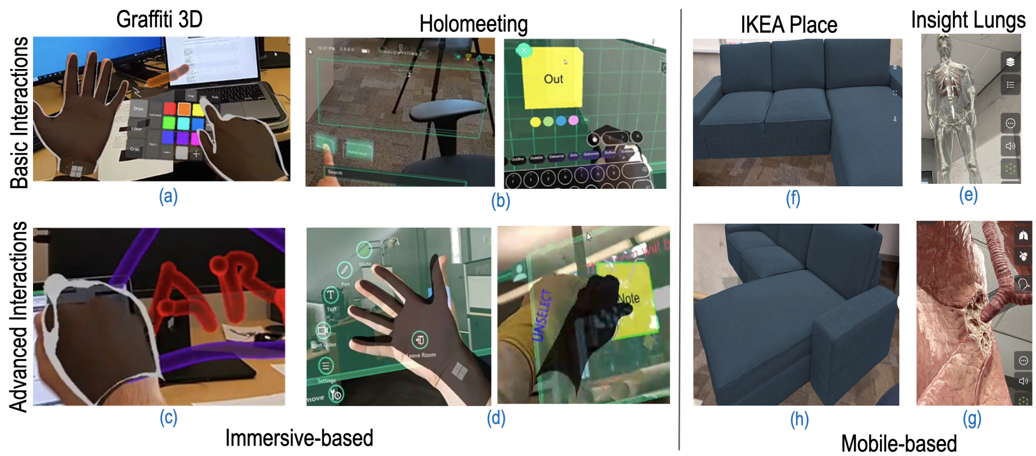

Fig. 1. Tasks and Interactions (Participant View): *Basic Interactions*: (a) selecting the draw option from a menu, (b) navigating to a room and adding a sticky note, and (e,f) calibrating virtual objects (e.g., a sofa or the human body); *Advanced Interactions*: (c) grabbing a sketch, (d) moving a sticky note, and (g,h) changing the size and seeing the structure of virtual objects from different angles.

[20, 87, 88, 95] or ensuring that the gestural interaction is comfortable [9, 64, 68, 74, 95, 99, 102]. An unresolved challenge for AR interaction remains; whether a new user can *discover the correct gestural input* required to use an AR system ([6, 42, 58, 86, 110, 114]) given the conceptual and cognitive interaction challenges novices may experience [62]. When novice users struggle or fail to discover the needed input on their own for initiating the interaction with the interface (i.e., gestures that are supported by a system), they are more likely to become frustrated and abandon their use of an application [34, 51, 75, 109, 114].

In this paper, we empirically study the first-time onboarding experience of novice users when interacting with immersive AR applications, specifically head-worn AR. To design input techniques that facilitate novice interaction with immersive AR, it is important to first understand how novices discover mid-air gestural interaction on their own and establish a better understanding of the interaction challenges they face while onboarding. The research questions guiding this exploration were:

- **RQ1:** How do novice users discover the gestural input required to interact with unfamiliar immersive AR applications?
- **RQ2:** What interaction challenges and barriers do novice users face while discovering gestural input in immersive AR environments?
- **RQ3:** What strategies or workarounds do novice users employ to tackle the challenges of discovering the needed gestural input in immersive AR?

To answer these questions, we ran an observational study with 25 participants from diverse backgrounds, occupations, and technical knowledge to deeply understand the challenges that “every day users” might encounter as this emerging technology becomes ubiquitous. This broad range of participants was selected to capture the most acute challenges that novice users may face [16, 85]. The inquiry focused on how participants initially discovered the gestural input needed to interact with unfamiliar immersive AR application, the types of interaction challenges that users experienced, and if and how they overcome these challenges and what are their workarounds to do so. For comparison, we also included non-immersive mobile AR applications that are more ubiquitous and available on phones and tablets that are more familiar to end users.

Our findings showed that the majority of participants largely struggled to discover the correct gestural inputs and relied on trial-and-error explorations. Although users were encouraged to watch tutorials and seek help *before* initiating their interactions, few participants used these resources, even after encountering numerous errors (consistent with the concept of the *active user paradox* observed in other research [22]). When participants did seek help, they struggled to locate relevant instructions and apply them to their context. A detailed analysis of gesture interaction traces in each task revealed that few errors results from system recognition issues and that the majority of errors occurred when participants could not determine the correct sequence of actions to perform (i.e., *gulf of execution*) or could not evaluate their actions (i.e., *gulf of evaluation*).

To contextualize the prevalence of our findings and their practical implications, we carried out an expert validation study with 8 professional AR designers, engineers, and researchers from 4 large technology companies. Reflecting on their real-world experiences, these experts confirmed that the discoverability issues revealed in our study are significant hurdles for nearly all newcomers to AR gestural interfaces. While the experts discussed potential improvements in low-level sensing and gesture recognition technologies to address these challenges, they collectively underscored the importance of accommodating users' diverse mental models and addressing user errors during the design phase. The experts further emphasized the importance of designing appropriate onboarding approaches that are natural and integrated into the application to facilitate better learnability of AR user interfaces.

The main contributions of this research are thus: (1) empirical insights into how users discover gestural input and interactions with AR applications and the importance of understanding interaction challenges to inform the design of AR interfaces; (2) a detailed analysis of interaction traces that revealed the key barriers impacting the discoverability of gestural input when interacting with both immersive and mobile AR applications (included for comparison); and, (3) insights from experts that confirm discoverability issues with gestural input and identify key opportunities for AR researchers and designers to improve the onboarding experiences of novices and help them form an accurate mental model of the interaction. Our detailed analysis and findings can be used to inform the design of input techniques for the next generation of AR interfaces, improve gesture recognition algorithms, and be applied to the creation of novel help systems for AR environments.

2 RELATED WORK

This research drew upon insights from prior work on user experiences with AR, the challenges incurred when designing and using gestural interaction in AR, and literature on how users learn to use unfamiliar feature-rich applications.

2.1 User Experiences with AR

AR emerged in in the early 2000's [67] and several design heuristics and principles have since been proposed [31, 33, 57, 59]. These have predominantly taken a technology-based perspective [41, 54, 83], focusing less on the user experience and the adoption barriers that exist [54, 85]. Studies of user expectations and experiences [29, 81, 82] have largely used survey-based techniques [83] and focused on niche scenarios (e.g., user expectations for enhancing shopping with mobile AR [82]). Most of the prior research capturing users' perspectives in AR is limited to mobile-based AR applications, resulting in few insights into the cognitive effects of immersive AR experiences on users [100, 113, 116]. Recent work by Woodward et al. [113], investigating children's use of different AR headsets, suggested the need to better understand the mental models of AR users so that inexperienced users can have a more seamless user experience. While these studies provide a high-level understanding of user experiences, they did not delve into the nuances of the gestural interactions that are needed to complete AR-specific tasks. Considering that AR leverages richer interaction modalities (e.g., mid-air gestures) that are unfamiliar to many users [12, 70, 83], there is a lack of understanding about how novice users discover gestural input with unfamiliar AR environments.

Research into user experience challenges in AR environments [54, 83] has largely considered the perspectives of users who are not representative of the average consumer. For example, Olsson et al. [83] considered perspective of users who were either experienced with mobile-based AR environments or had advanced educations (e.g., academics with Computer Science training or Hololens users). While this research provided valuable insights, the audience of AR applications today is no longer limited to technical experts (e.g., Dey et al. [28]). Thus, our present research explores how novice users from diverse educational and professional backgrounds, who are more representative of the current audience of future AR products, would discover gestural interactions with AR applications.

2.2 Challenges When Designing and Using Gestural Interactions

Research into AR interfaces and interactions has a rich history (see survey by Billinghurst et al. [17]). In our present research, we focus on the mid-air gestural interactions currently employed within immersive consumer AR devices [27] and touch-based interactions in mobile AR devices. The use of these modalities can be traced back to early gesture research, such as gestural elicitation [58, 89, 107] (both user-led and expert-led) across different platforms, including interactive tabletops [71, 111] and mobile [36]. Despite the extensive research on gesture elicitation, there are only a few studies that focus on the gaps between users expectations and proposed gesture sets [71] or highlight the importance of additional factors such as gesture-action mappings that can facilitate the adoption of various gestures [49]. Some efforts have also been made to better understand users' perceptions and behaviors during gestural interaction in spatial environments, such as AR [54, 113], including mid-air gesture elicitation [6, 58, 86, 108, 110] to understand users' behaviours and expectations while manipulating virtual objects in immersive AR (e.g., Microsoft Hololens 1 [86], Magic Leap One [110], Wizard of Oz [58], a video see-through AR setup using an HTC Vive headset [6], etc.). However, the primary focus has been on informing the design of discoverable and memorable gestures for different AR environments for specific target user groups. In contrast, we focus on how novice, everyday consumers who are unfamiliar with AR discover gestural input and the types of barriers they face while discovering and learning gestures in commercial AR applications.

Research in AR has, so far, acknowledged issues and breakdowns related to technical challenges [7, 15], precision issues [20, 87, 88], system errors [7, 114], social acceptance challenges [7, 15, 63], privacy concerns [50], and fatigue while using gestural interaction [9, 64, 68, 74, 99, 102]. Extensive research has explored efficient interaction techniques [20], and devised alternative ways of interacting in AR (e.g., using a user's body as an input device [63, 72, 95, 115], using external devices [24, 112, 117], etc.) to resolve precision issues [20, 87, 88, 95] and fatigue problems [9, 64, 68, 74, 95, 99, 102]. However, issues related to the learnability of gestures in AR [14, 21, 114]) and the cognitive impact of user errors [45, 62] are only starting to be recognized. In fact, Norman and Nielsen [77] argue that gestural interfaces are a step "backward in usability" as fundamental interaction design principles are often not considered. Considering AR applications are a tremendous leap from the desktop and mobile-based interfaces that most users are familiar with [17], understanding user experiences and interaction challenges and designing discoverable interaction techniques is necessary to ensure the long-term adoption and usefulness of these applications [56]. The present study complements the existing research by conducting an in-depth error analysis of gestural interaction and by examining novice users' gulfs of execution and evaluation and other proposed guidelines for gestural interactions [77]. We contribute new knowledge about how novice users' experiences with AR applications led them to discover gestural interactions and identified when they encountered breakdowns during different task-specific gestural interactions.

2.3 Learning and Onboarding Within Complex Applications

One goal of our research was to understand how novice users approach discovering and learning the needed gestural interactions when using AR applications and users' workarounds when they encounter any barriers. To contextualize the findings, we consulted a range of prior research that has investigated the challenges inherent in learning complex feature-rich applications across 2D and 3D domains and users' help-seeking strategies for resolving breakdowns [40, 44, 52, 55, 66, 94]. For example, such studies have shown that users prefer self-directed experimentation or trial-and-error activities within an application rather than having to seek out instructions and help [55, 93, 94]. This phenomenon is known as the *active user paradox* [22]. Some studies have also found that users were reluctant to leave an application or task context to seek help from external resources [78–80]. Some studies [55] have also shown that users who leave an application to look for help within an online system often become trapped in the *vocabulary problem* [39], i.e., they struggle to use the correct terminology when looking for help.

Inspired by these empirical insights, help systems have evolved from formal documentation and manuals [93] to contextual help systems embedded within applications [19, 25, 43, 46, 60, 105]. While there is a rich history of work supporting software learnability and easing help-seeking processes [25, 37, 43, 61] and recent work has explored the challenges for user onboarding in VR [23], it is unclear whether users' learning approaches and strategies in 2D/3D software domains or VR apply to AR, which has distinct interaction challenges caused by 3D virtual objects within real-world environments [41]. The present research extends the prior work by examining if and how novice users approach discovering and learning the needed gestural interactions, in AR applications and what are their workarounds in overcoming interaction challenges. The study also sheds light on the type of assistance that different users expect to receive when resolving such breakdowns.

3 METHOD: OBSERVATIONAL USER STUDY AND INTERVIEWS

We ran an in-lab observational user study with 25 participants from various backgrounds who were all new to AR and even virtual reality (VR). The goal was to understand how novice users discover appropriate gestural input techniques across immersive modalities and non-immersive mobile AR (included for comparison), how users recognize and articulate any potential barriers they encounter, and which strategies they use to resolve such barriers through help-seeking.

3.1 Participants

We recruited a diverse participant pool representative of everyday consumers and technology adopters: little to no prior experience with AR or VR, and an interest in using these technologies in the future. Participants were recruited via advertising posters at local educational organizations and reaching out to the organization's administrative and support staff in person and through their mailing lists. Further, we recruited 4 additional participants who did not have experience with AR, but had prior experience with other forms of 3D interaction (e.g., 3D design or playing VR games) from a local maker space and through snowball sampling.

The participant pool included 25 novice AR participants (14F,11M) covering a range of age groups: 18-24 (29%), 25-34 (58%), 45-54 (4%) and over 55 (9%). The participants represented a diverse group of working professionals, including office administrative staff (7/25), academic researchers - Computer Science (3/25)/ non-Computer Science (3/25), warehouse workers (3/25), software/machine learning engineers (3/25), 3D printing experts (2/25), information designers (2/25), a

PID#, Gender	Occupation	Age	Background and Major
P1 (M)	Researcher	25-34	Information Science (MSc)
P2 (M)	Software Engineer	25-34	Computer Science (PhD)
P3 (F)	Admin Assistant	25-34	Educational Technology (BA)
P4 (F)	PhD Student	25-34	Computer Science, Robotics (PhD)
P5 (F)	Maker Space Assistant	45-54	Kinesiology (BSc)
P6 (M)	Program Assistant	25-34	History (BA)
P7 (F)*	Program Assistant	25-34	History (BA)
P8 (M)	Student	18-24	Post-Secondary
P9 (F)	Student	18-24	Computer Science, Computer Vision (MSc)
P10 (M)	Machine Learning Engineer	25-34	Computer Science (MSc)
P11 (F)*	Office Admin	18-24	Post-Secondary
P12 (M)	Student	18-24	Engineering Science (BEng)
P13 (F)	Student	25-34	Geography (MSc)
P14 (M)	Software Engineer	25-34	Computer Science (MSc)
P15 (F)	Maker Space Assistant	25-34	Interactive Arts (MSc)
P16 (F)	Student	25-34	Psychology (MSc)
P17 (F)	Logistics Supervisor	18-24	General Studies Diploma
P18 (M)*	Construction Worker	25-34	Marketing Diploma
P19 (F)	Custodian	25-34	Post-Secondary
P20 (F)	Warehouse Worker	18-24	Post-Secondary
P21 (M)*	Warehouse Worker	18-24	Post-Secondary
P22 (F)	Office Admin	55-65	Chemistry (BSc)
P23 (F)	Grant Coordinator	55-65	Biology (BSc)
P24 (M)	Post-Doc in Visualization	25-34	Visualization (PhD)
P25 (M)	Financial Coordinator	35-44	Business (MA)

Fig. 2. Summary of Participants (* None of the participants had any experience with AR environments, i.e., mobile or immersive headset except for a few participants (4/25) who had briefly used VR games or 3D design)

custodian (1/25), and a sales representative (1/25). Participants also had different levels of education, ranging from high school degrees to graduate degrees in various academic backgrounds (Figure 2).

3.2 Apparatus and In-Lab Setup

For our study, we used the head-mounted display Microsoft Hololens as it offered a more complete ecosystem of consumer-grade, non-gaming related immersive AR applications at the time of writing this paper [32]. The immersive AR applications that we tested were hosted on a Microsoft Hololens 2 with 2k resolution and >2.5k radiants (light points per radian) Holographic density. We used mobile-based AR applications hosted on a Google Pixel 6 Android phone (6.2” X 2.9” X 0.4” with a 1080 x 2400 OLED resolution and 207 grams weight) and an iPad mini (7.69” X 5.3” X 0.25” and 297 grams weight with a 2266 X 1488 resolution).

The study was conducted in a spacious lab setting that enabled participants to freely explore AR objects alongside real world objects. Multiple cameras in the environment captured participants’ gestures and interactions during the study. Sessions and participants’ think-aloud commentaries were video and audio-recorded for later transcription. In addition, we captured screen recordings of the AR applications and there was at least one researcher present in the room to observe each session and take notes.

3.3 Choice of AR Applications and Tasks

To select tasks and AR applications for the observational study that would be appropriate for a diverse set of consumers, we surveyed different *(consumer-grade task-oriented AR applications)* in domains such as maps and directions, medical anatomy, shopping, education, and virtual collaborations and meetings. To better motivate diverse set of novice users to perform tasks with AR applications during the study, we focused on popular productivity-related or day-to-day tasks that everyday consumers could find relevant and we avoided niche gaming applications.

We explored applications that were already popular on the Hololens 2 platform [92] for boosting productivity or performing day-to-day tasks related to education, remote collaboration, shopping, among others. We also considered applications and tasks that would require different interaction techniques (e.g., voice, typing, gestures, touch interactions) that could lead to different types of AR breakdowns.

After our initial exploration, we selected 4 popular task-based AR applications that we deemed suitable for everyday users. For immersive applications (i.e., on the Hololens 2), we used (i) *Graffiti 3D* for drawing in 3D and (ii) *Holomeeting* for remote collaboration(Figure 1). For the mobile scenarios, we used (iii) *Insight Lungs* (an Android application on the Google Pixel 6) for learning about medical anatomy and (iv) *IKEA Place* (an ioS application on the iPad mini) for virtual furniture placement in real space. For each application, we designed tasks that would require a range of interactions and allow participants to have a richer interaction with virtual AR objects and the real-world environment.

For each application, we considered interactions involving both basic and advanced gestures. For the *basic interactions*, participants would need to perform unimanual gestures [108] or use voice commands for tasks such as selection, typing (immersive-based AR), creating a virtual 3D sketch on real-world object such as laptop (immersive-based AR, Figure 1c) or to perform single-touch interactions (mobile-based AR). For the *advanced interactions*, participants would need to perform bimanual gestures [108] to access menus or grab and rotate objects (immersive-based AR) and multi-touch interactions to perform tasks such as rotating or resizing objects (mobile-based AR). For example, one of our *basic interaction* tasks with Holomeeting asked participants to search for a meeting room in an AR immersive application. For the *advanced interaction*, they used the whiteboard in the meeting room (Figure 1b) to add sticky notes, and moved the sticky note with respect to their real-world surroundings (Figure 1d).

Additionally, we included a few revisiting task interactions that required reapplying gestures participants discovered and learned across applications (e.g., Selection, Grabbing, Getting a menu, etc.) to gain a better understanding of how participants would apply an interaction once they were familiar with the underlying application, technology and the gestural interactions.

3.4 Study Design and Procedure

We used a within-subject design to study different errors and how users recover from them while minimizing the effect of inter-participant variability. Each participant used the four AR applications, starting first with the interactions using basic gestures and then again using the advanced gestures.

3.4.1 Introduction to AR and Training. Each study session began by introducing the participant to the concept of AR and provided some general tips to interact with the application (e.g., using their hands). We provided participants with a laptop with web access and encouraged them to seek out any online tutorials or help resources within the Hololens device (e.g., [97], [1]) before or while interacting with any AR application. We did not force participants to watch a specific tutorial as we were interested in understanding how users discovered gestural interactions *on their own* during their use of an AR application and what strategies they used when they encountered challenges. The researcher who was in the room did not offer any help once the study was in progress. In the worst case scenario when a user was completely stalled, the researcher offered some tips. The AR applications were already setup and we did not ask participants to do any other set up (or watch any Hololens setup tutorial [98]) as we were not observing their interactions while setting up AR devices. Furthermore, the tasks that we observed required more sophisticated gestural interactions not offered in the setup tutorials. For both mobile and immersive AR applications, the users were handed the devices with the applications already open so they could focus on the given tasks. In the case of immersive AR, the researcher helped the user to wear the head-mounted display with the application open.

3.4.2 Study Protocol. Participants first completed a demographic questionnaire on their background and prior experiences with AR/VR applications. Next, each of the AR applications was presented to the participant, one by one, in a random order. Each participant completed two tasks (with basic and advanced interactions) for each application on

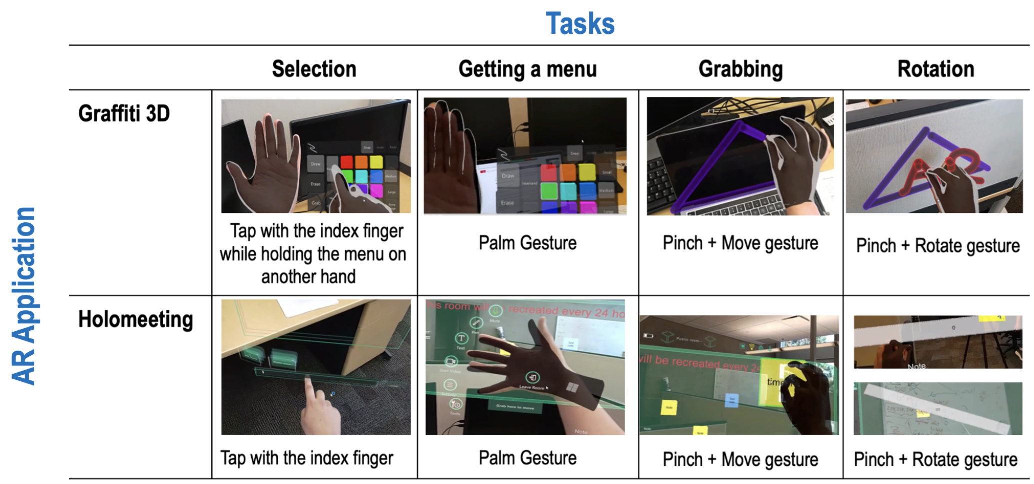

Fig. 3. Gesture analysis: Description of correct gestures expected from users to perform given tasks in immersive AR applications

two different devices (Immersive and mobile) in a random order. We used a Latin Square [91] to balance the order in which tasks were presented. We provided 15 minutes for participants to explore each immersive AR application, and 7-8 minutes with each mobile-based application. An additional 5 minutes was allotted to complete the post-task questionnaire (hosted on the laptop provided) after using the basic or advanced interactions with each application. We encouraged participants to think aloud and reminded participants that the study was seeking to understand how they interacted with, and used, AR applications rather than their performance or ability to master AR applications or use the Hololens device.

Lastly, we conducted follow-up interviews to further probe any difficulties that impacted the use of each application, any AR features or interactions that worked or did not work well while using the applications, and any consequent help-seeking attempts that were made while using each application. Each session lasted approximately one hour and participants were provided with a $15 CAD Amazon gift card in recognition of their time.

3.5 Data Analysis

To investigate the mental models that participants constructed and understand how participants discovered the different gestures that were needed to interact with AR, we coded and analyzed the video, audio, and screen recordings using an inductive analysis approach [26]. We used Norman's action cycle and fundamental cognitive principles of discoverability and interaction [75] to identify initial themes relating to how participants discovered the needed gestures in AR applications, and the types of errors and interaction challenges they experienced. We further analyzed the user interactions in each AR application using fundamental Norman-Nielsen (NN) principles of interaction design relevant to gestural interfaces [77], shown in Figures 9, and 10. Through discussion with the rest of the research team and the use of physical affinity diagrams [26], we identified initial insights about participants' struggles and the core themes evolved into a single coding scheme.

Gesture Analysis: To analyze how participants initially discovered the correct gestures to use, we coded the video recordings and screen recordings from each AR application and analyzed each participant's gesture trial attempts for major input task actions (i.e., Selection, Typing, Getting a menu, Grabbing, Rotating). We looked at first three gesture trials of participants for each task action and analyzed to understand the occurrence of different error themes among

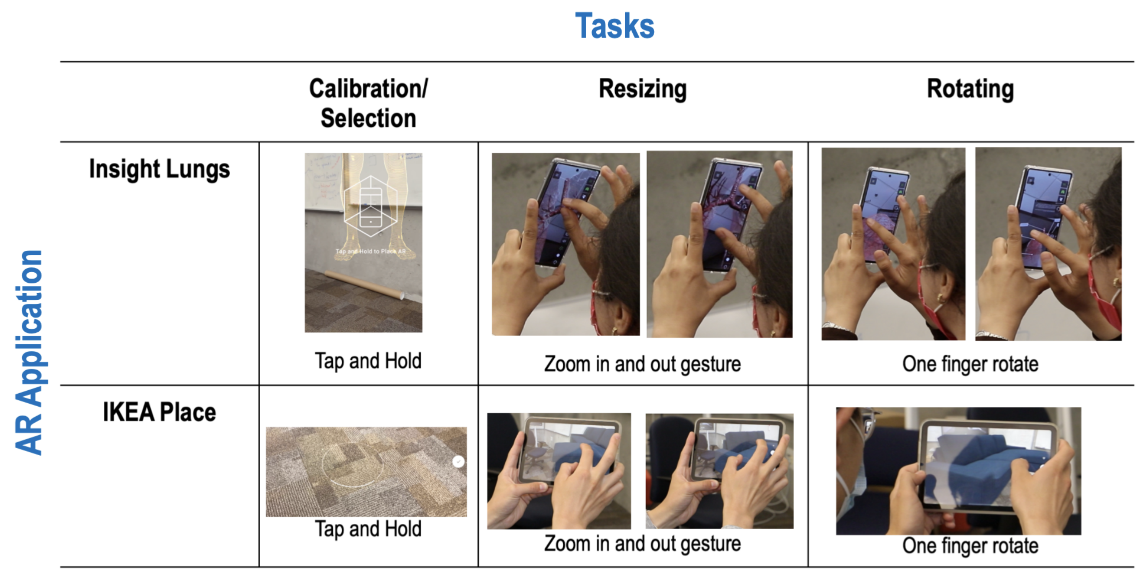

Fig. 4. Gesture analysis: Description of correct gestures expected from users to perform given tasks in mobile AR applications

the total unsuccessful trial attempts across all participants. We organized participants gesture trials into a 2x2 confusion matrix [35, 90, 103, 104] with two dimensions ("gesture" and "outcome") to demonstrate participants' error trial attempts and how they discovered the gesture. Lastly, we corroborated the data with participants' think-aloud verbalizations to identify the factors that impacted the discoverability of these gestures.

Help-Seeking Analysis: To investigate how participants tried to discover the correct gestures to perform, we identified the point at which participants decided to seek help (e.g., by searching the web or asking the experimenter [55]) using each participants' search queries and browser-based navigation history, their actions from the screen recordings, and their think-aloud verbalizations.

3.6 Presentation of Results

We have organized results according to the key themes answering our research questions: 1) how participants discovered the gestural input required to interact with immersive AR applications and mobile AR applications (included for comparison); 2) the types of errors and interaction challenges that participants faced during their discovery of gestural input with unfamiliar AR apps; and, 3) the types of barriers participants faced when they tried to seek help to learn the proper interaction.

4 DISCOVERING GESTURAL INPUT

To gain an in-depth understanding of how users discovered gestural input, we conducted a detailed analysis of mid-air gestural interaction *(immersive AR)* and, for comparison, also analysed the touch-based interactions *(mobile AR)*. To contextualize the observations, we used the paradigms of *gulf of execution* (e.g., gaps when determining the correct sequence of actions), *gulf of evaluation* (e.g., gaps when evaluating the effect of an action) [75] and fundamental NN principles of interaction design relevant to gestural interfaces [77].

Selection (Holomeeting); Total 62 trial attempt among all participants

	Correct Gesture	Wrong Gesture
Correct Outcome	True Positive(24%) 24%; (12/25): Tap with index finger	False Positive (10%) • 8% (4/25): Tap with two fingers (index + middle)* • 2% (1/25): Pinch wrapping around the button
Wrong Outcome	False Negative (18%) 18% (2/25): System Error=Depth perception (cursor)/recognition error	True Negative (48%) • 26% (7/25): Tap with two fingers (index + middle)* • 16% (3/25): Tap on headset

Getting menu (Graffiti): Total 69 trial attempts among all participants

	Correct Gesture	Wrong Gesture
Correct Outcome	True Positive (10.5%) 10.5% ; (6/25): Palm Gesture	False Positive (35.1%) • 12.3% (6/25): Palm oriented sideways [90 degree] • 10.5% (4/25): Fist both hands • 8.8% (5/25): Palm both hands*
Wrong Outcome	False Negative (12.3%) 12.3%; (7/25): System Error-Depth perception (cursor)/recognition error	True Negative (42.1%) • 22.8% (9/25): Back palm • 12.3% (6/25): Others (Pinch*/ pointing with index finger + move) • 7% (4/25): Palm both hands*

Grabbing (Holomeeting): Total 68 trial attempts among all participants

	Correct Gesture	Wrong Gesture
Correct Outcome	True Positive (8.9%) 8.9% ; (5/25): Pinch + move gesture	False Positive (23.2%) • 16.1% (7/25): Fist + move/lift (one/ both hand)* • 7.2% (4/25): pinch with two or more + move/drag (one/both hands)
Wrong Outcome	False Negative (7.1%) 7.1%; (4/25): System Error-Depth perception (cursor)/recognition error	True Negative (60.7%) • 26.8% (11/25): Fist + move/lift (one/both hand)* • 21.4% (8/25): Palm + move • 5.4% (2/25): Physical orientation (head, body)

Rotation (Holomeeting): Total 41 trial attempts among all participants

	Correct Gesture	Wrong Gesture
Correct Outcome	True Positive (10.3%) 10.3%; (3/25): Pinch + rotate	False Positive (10.3%) • 6.9% (2/25): Pinch with two or all fingers +move* • 3.4% (1/25): Fist + move, physically stood up and went closer
Wrong Outcome	False Negative (6.9%) 6.9% (2/25): System Error-Recognition error/Calibration line or depth perception	True Negative (72.4%) • 31% (6/25): Fist + move • 24.1% (7/25): Pinch with two or all fingers +move* • 17.2% (3/25): tap/pointing with two/more or index fingers

Fig. 5. Analysis of the frequency of errors during mid-air gestural interactions. Each table summarizes the observed errors for one of four gestures (i.e., Selection, Getting a menu, Grabbing, and Rotation), organized as a confusion matrix. The percentages are the trial attempts for each error gesture instance and the ratio is the number of unique participants who attempted that erroneous gesture. System errors included false positive and negatives (e.g., at least 7/25 participants experienced false positives during grabbing). True negatives overall dominated the errors across all tasks and represented human errors and misconceptions (e.g., for the selection task, 7/25 participants wrongly tapped with two fingers across 26% of total 62 trial attempts). *denotes gestures that worked ambiguously for some participants but not for all participants, resulting in false positives for some and true negatives for others.

Consistent with behaviours of software users in other contexts [55, 80], the first instinct of most participants (23/25) was to explore and figure out the AR interface on their own, rather than watch tutorials or seek help. As one participant explained: *"I start with trying out myself and try to guess how it will work..." (P20).* This demonstrated the *active user paradox* [22]: trial-and-error behaviors dominated the gestural interactions, especially with immersive AR apps. In fact, the majority of participants (15/25) could not formulate clear intentions in their minds [75] about the immersive interface nor mid-air gestural interactions, thus lacking a mental model about how to use gestural input to interact with immersive AR. Unlike mobile AR touch-based interactions, mid-air gesture interaction was not intuitive: *"I think mobile/iPad, are much easier and familiar. In the [Hololens]...I was not familiar with what my hands or actions could do... I was struggling, I was figuring, I was exploring, imagining if this could work or not"* (P17).

Some participants (10/25) discovered gestural interaction by accident or in some cases by leveraging the visual affordances and feedback offered by the application. For example, there was accidental invocation of the menu (6/25 participants): *[Graffiti 3D]: "...it was a fluke of getting that palette so, I do not know what action I did"* (P14). Another participant, even after being aware of the interface in VR space, could not map the affordance provided by the interface: *[Holomeeting]: "I was trying to grab, it did not grab, and [then] I accidentally opened my hand...and [the] menu appeared"* (P18). We observed that even users who had experience with VR environments or 3D interfaces eventually struggled the same way as other users with misleading signifiers, confusing mappings, and trial-and-error loops. Another cue that assisted 9/25 participants during their use of *(Graffiti 3D)* was the visual outline of their hands: *[For getting a menu]: "I see some shadow of my hand and tried random gestures and, now I see different colors"* (P16). However, a few participants (5/25) were confused about what could be mapped to this visual affordance: *"I can only see the outline of my hand and little globes appearing but I do not know what to do with it."* (P3)

In contrast, with mobile AR apps, most participants (19/25) had an easier time as the mobile UIs were consistent with non-AR apps and provided visibility of system status. Since they were already familiar with touch-based gestural input, participants were also enthusiastic and excited about trying out AR, as seen in other user studies [83].

Overall, confusion at the outset about *how to discover correct gestural input with immersive AR* significantly impacted the remaining user interactions and eventually led to different types of errors, explained in the next section 5.

5 DETAILED ERROR ANALYSIS OF INITIAL GESTURAL INPUT

When using the mobile AR applications, the key errors that we observed were related to calibration and recognition issues (i.e., *system errors*; see Section 5.2). These errors were mainly due to misleading signifiers and violation of reliability design principle leading to incorrectly learned interactions (See Figure 9 and 10). In contrast, both the immersive AR apps violated several NN design principles [77], making discoverability of gestural interactions more challenging (See Figure 9 and 10). As a result, we found high rates of both system and human errors with immersive AR applications, mainly due to users lacking an accurate mental model of mid-air gestural interaction. We conducted a detailed gesture analysis for each Hololens task and highlighted clusters of errors observed during the first 3 trials. These results are presented as a confusion matrix with two dimensions ("gesture" and "outcome"). We also present observations related to users' mental models and how these different errors impacted the discoverability of gestures.

5.1 Overview of the Gesture Analysis

Based on the criteria presented in Section 3.5, we analyzed the first three gesture trials of each participants for each task that required a range of different gestures (Figure 3, 4). We considered basic interaction involving unimanual gestures (e.g., *Selection*, *Typing*, *Sketching* tasks) and advanced interaction involving bimanual gestures (e.g., *Getting a Menu*, *Grabbing* and *Rotation* tasks). We characterized participants' gesture trials in the confusion matrix as follows: (1) *True positive:* correct gesture and correct outcome, (2) *False positive:* wrong gesture but ambiguously correct outcome, (3) *False negative:* correct gesture but wrong outcome due to system errors such as recognition error, depth perception error or even lack of system feedback, and (4) *True negative:* wrong gesture and wrong outcome showing range of wrong gestures tried by participants. For the error analysis, the errors obtained in immersive AR were grouped into two categories: (i) System-induced errors involving false negatives and false positives and (ii) Human errors involving true negatives.

Overall, nearly half of the participants (12/25) could successfully discover the correct basic gestures (e.g., selection) within the first 3 trials, but far fewer participants (6/25) were successful with the advanced interactions (e.g., grabbing). Users often faced difficulties due to gulf of execution (e.g., gaps in figuring out the correct sequence of actions), but for more advanced interactions, issues also occurred due to gulf of evaluation (e.g., gaps in evaluating the effect of an action). We observed different system and user errors that impacted users' mental models (explained in Section 5.2, 5.3 in detail). Interestingly, true negatives (Figure 5) were the leading cause of errors, followed by false negatives with the basic interactions and false positives with the advanced interactions (see Section 5.2). A detailed explanation of the confusion matrices (Figure 5) for each task in immersive AR is in the Appendix (see Appendix).

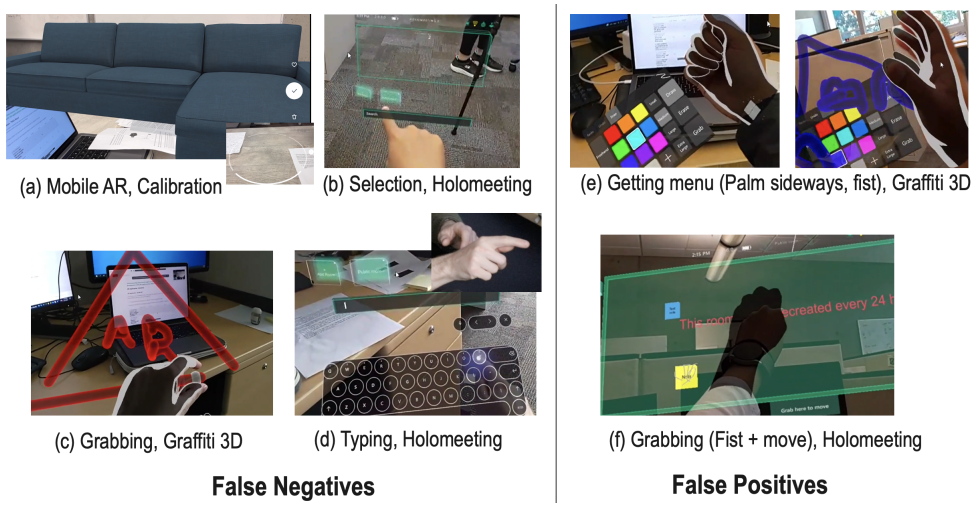

Fig. 6. System Errors: Instances of false negatives and false positives from all participants (explained in Section 5.2)

5.2 The Impact of System-Induced Errors

Next, we describe the impact of each error type (e.g., system-induced errors: false positive and false negative; human errors: true negatives) in detail.

False positives: In comparison to mobile-based AR, the most common system errors in immersive AR were due to false positives and had a higher impact on the discoverability of gestures. On average, out of a total 398 trial attempts (across all 8 tasks, and 25 participants), false positives occurred 14.8% of the time and were encountered by about half of the participants (12/25) at least once. Among the different tasks, the *Getting a menu* task with Graffiti 3D accounted for the highest number of false positives (35.1% out of total 69 attempts; 12/25 participants).

When participants encountered false positive errors, they were confused about how things worked previously and backtracked to recall what gesture worked for them earlier (Figure 6.e), e.g., *[Getting menu, Graffiti 3D]: "I need to try get that [menu] again, but how did I do that before?"* (P13). However, due to an ambiguous mapping of the gesture to the correct outcome, users kept trying other gestures, eventually developing an incorrect mental model of the interaction. Interestingly, participants who experienced false positives attempted even more trials (average: 5 trials, range: 1-8 trials) than those who did not experience false positives (average: 3 trials, range: 1-9 trials) for the *Getting a menu* task with Graffiti 3D.

Due to the lack of consistency in the outcomes produced by false positives, participants formed some misconceptions (Figure 6f). For example: *[Grabbing, Holomeeting]: "I do not know what I did, but I moved the outline. I think I did something that is how I was able to move, maybe used both fingers? [wrong discoverability of gesture]"* (P17). Eventually, most users failed to discover the correct gesture within 3 trials: *[Getting a menu, Graffiti 3D]: "I thought I figured out how to interact with an object but it was not consistent...I saw that [menu] comes if I do this [fist both hands], but when I wanted to do that again, it did not work."* (P15). A detailed analysis of P13, who had the highest number of false positives during their maximum trials (Figure 7b) demonstrated the impact of discoverability when participants kept on backtracking to determine which wrong gesture worked during *Getting a menu* task in Graffiti 3D.

False negatives: In immersive AR, false negatives constituted one-third of the errors. On average, 11.1% of total 398 trial attempts (across all 8 tasks, and 25 participants) were false negatives and were experienced by 10/25 participants at least once. These system-related errors were caused by issues with recognition or depth perception, and/or a lack of system feedback. With false negatives, most participants failed to recognize that they discovered the correct gesture and performed more random trials. As noted by P21 (even after having experience with VR or 3D interfaces) and P15, they were confused even after attempting the correct gesture due to false negatives (Figure 6c).

Compared to immersive AR, false negative errors dominated mobile AR tasks when the application failed to recognize real-world objects. This resulted in calibration and recognition issues, as noted in other mobile AR studies [15, 45]. Many participants (15/25) struggled to determine the meaning of AR features and followed interactions which did not work as they expected, resulting in frustration (Figure 6a). For example, P11 even after having prior experience with 3D and VR games, struggled with misleading signifiers and their corresponding mapping: *"It [IKEA Place] is not placing the sofa in the correct location, it is floating in the air. I worked a little bit with that [calibration ring] to see if I can make the sofa a bit lower down...which is frustrating, but it is not on the floor."* (P11). Even after several attempts (e.g., rotating the phone left/right, physically moving themselves to space, following AR feedback of pointing to vacant space, etc.), participants were unclear why these errors occurred.

Next, we shed light on the detailed impact of false negatives on the discoverability of mid-air gestural interactions in immersive AR using the case study of P7 (Figure 7a), who even after having prior expertise in VR environments and 3D spaces, experienced the highest number of false negatives with the *Selection* task in Holomeeting (task having the highest occurrence of false negative trials: 18% of total 62 attempts). P7 encountered false negatives in their earlier 5 attempts, failing to recognize the correct gesture they had already performed. This led to an additional three loops of random trial-and-error (denoted by *), with P7 ultimately seeking help from external resources (arrows highlighted in black) and the researcher to finally discovering the correct gesture. These results underscore the struggle users experienced in AR environments, which is different from other VR or 3D interfaces and emphasize that the prior knowledge of VR did not necessarily result in a positive transfer to AR.

5.3 How and Why Human Errors Dominated the Interaction

Although the system errors were significant in immersive AR, human errors (i.e., true negatives) dominated problems with user interaction. In contrast, in mobile AR, we observed true negatives only among a few participants (8/25) who struggled with AR features and interactions while calibrating the virtual objects with real-world surroundings. In immersive AR scenarios, human errors were surprisingly higher and accounted for two-thirds of all errors across all tasks (i.e., an average 63.5% of total 398 attempts experienced by 25/25 participants within the first three trials). True negatives in immersive AR were induced by wrong gestures that participants expected to produce a correct outcome. In immersive AR, we observed that participants' mental models were affected by a range of misconceptions and/or the use of other types of technologies, which we categorized into three broad categories.

5.3.1 Transferring real-world object interaction to virtual objects. Across all tasks, about half of the participants (12/25) tried performing gestural interactions with virtual objects similar to how they would when using real-world objects. For example, in grabbing and rotation, participants *"would grab normally"* (Figure 8a) and expected the gesture to work using fist and move. Similarly, 3/25 participants mistakenly perceived that as the virtual object (e.g., screen, sticky note in Holomeeting) is seen through the headset and/or headset lens, they would need to physically place the sticky note on the headset glasses or tap on the headset (Figure 8b,d).

Some participants (6/25) assumed that physically moving themselves closer to a virtual object (e.g., keyboard, virtual screen in Holomeeting) while performing the gesture (e.g., selection, grabbing and rotating task action) was needed

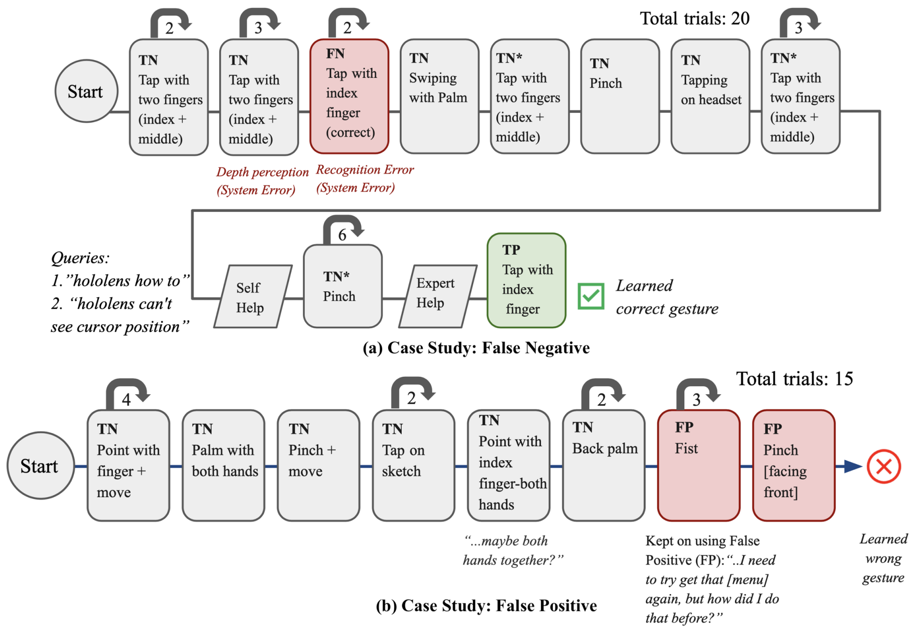

Fig. 7. Case study of false negatives and false positives impacting discoverability: (a) With false negatives, participant (P7) failed to recognize that they discovered the correct gesture and performed more random trials (during *Selection* task in Holomeeting) (b) With false positives, participant (P13) kept on backtracking to determine which wrong gesture worked and eventually learned the wrong gesture (during *Getting a menu* task in Graffiti 3D). In (a) Case Study: False Negatives, * denotes the gesture users already attempted but failed to recognize them as correct gesture. Therefore, they performed additional three loops of trial-and-error.

because the virtual object was not at the correct distance from them (Figure 8c), e.g., *"...the keyboard is way too close so I have to stay backwards and trying to adjust the position of the keyboard"* (P11). Such proximal gestures have also been described in the literature [86].

5.3.2 Confusing Mappings: Misinterpretations leading to inaccurate explanations of how AR application and interactions works. Some participants (8/25) experienced a gulf of execution when mappings in the immersive UI were confusing or unclear. For example, in the selection and grabbing tasks, participants held assumptions about the mapping of the line coming out of their hands for correctly selecting the menu items or even tracking their interaction (Figure 8f) (also termed as distal gesture [86]). Similar errors occurred when participants started mapping their head movements to move virtual objects or coordinated their head and hand movements to access virtual objects. For example, P13 said, *"I have a difficulty in accessing the menu [as] I think my head's movement and hand is not [happening] at the same time. They are not covering each other..."*. A few participants were confused about which hand to use of perform the correct gesture (e.g., Selection, Getting a menu tasks; Figure 8e). For example, *"for any reason, it is all always scan my left hand instead of the right hand."* (P8).

5.3.3 Misaligned assumptions about virtual interactions (observed across all tasks). Although many participants (13/25) had the correct intuition about which gesture to use, they experienced errors due to incorrect execution. For example, 6/25 participants made false assumptions about the position, orientation, or pressure needed for the correct placement of the gesture with respect to a virtual object (Figure 8g,h): *[Selection, Holomeeting]: "I have to do a little bit this [tapping gently]. Sometimes a little pressure was helping."* (P17). In other cases, 7/25 participants made other assumptions: *[Selection,*

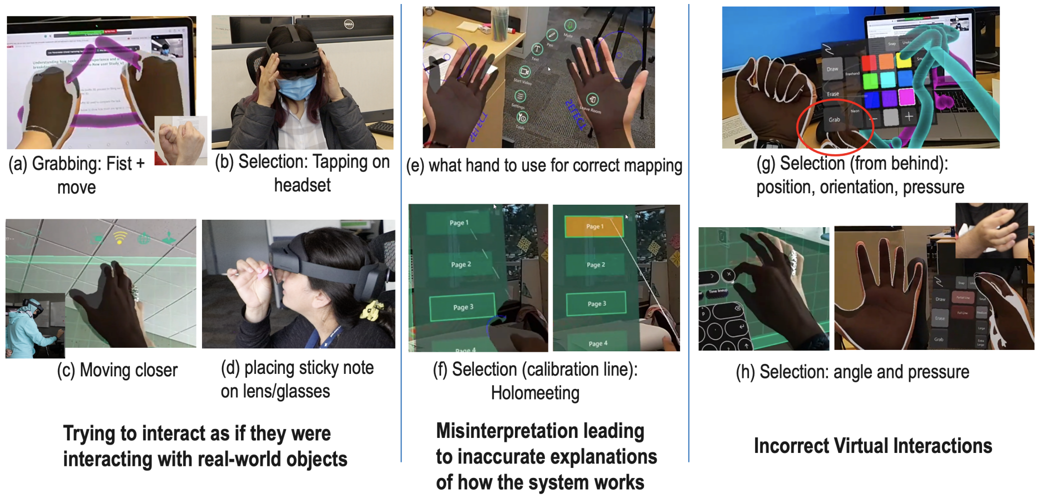

Fig. 8. Human Errors dominated the interaction: Instances of true negatives from all participants in immersive AR

Holomeeting]: "I think the angle of my hand matters"(P9). The variation in participant perceptions and their lack of an accurate mental model for executing various tasks led to different human errors, hence the increased true negatives. Some participants, however, could eventually determine the correct gesture in 3 trials as well as maximum trials (beyond the initial 3 trials). In the next section, we describe some efforts and ways in which users tried to discover the correct gesture by making attempts to discover the correct actions.

6 CUES AND WORKAROUNDS FOR DISCOVERING GESTURAL INPUT IN AR

Although several participants (19/25) accidentally stumbled upon the correct gesture at least once during their trial-and-error process, participants also tried to leverage the visual cues or feedback provided by the application. They could sometimes find workarounds by transferring their learning from previous tasks, or seeking help online. Next, we discuss these behaviors and efforts and highlight how participants still struggled to form the correct gestures to complete the tasks.

6.1 Visual Cues and Feedback Provided by Applications

Across all tasks, almost half of the participants (11/25) noted that the visual cues and feedback provided by the immersive AR application (e.g., outlining users' hands, cursor ring on 3D objects) helped "kick-start" their gesture learning. However, participants felt that this feedback was inadequate, inconsistent, and the mappings were not obvious. For example, P3 shared their frustration with a visual cue in the Graffiti 3D Grabbing task: *"I see that prompt [cursor ring around the 3D objects]... I thought that maybe the cursor was the tip of my finger, and you can change the orientation but I do not know how it works"* (P3). Some participants (7/25) also expressed a need for more in-application instructions about what gesture to perform. For example, after P9 learned the gesture to find the menus, they noted out that *"it was not written anywhere or instructed anywhere in the application"*. P2 further expressed a desire for *"onboarding or help experience that would really give me that kind of information I needed about how to use it [the app]...a lot of it just kind of assumed that I would figure out the gestures myself..."*.

	NN Principle: Visibility of system status (Non- existent signifiers)		NN Principle: Visibility of System status (Misleading signifiers)	
Application Name	**Y/N**	**Evidence showing interface followed or violated NN principle**	**Y/N**	**Evidence showing interface followed or violated NN principle**
Graffiti 3D	⚠️	*Visual affordances* (e.g., hand outline, cursor ring) but users didn't know what could be mapped	❌	*Misled affordances:* that did not work (e.g., affordance of AR headset/lens) Mappings not obvious (e.g., b/w cursor ring & orientation)
Holomeeting	❌	*No visual affordance/* feedback (e.g., while grabbing, tap on the keyboard) due to depth perception	❌	*Misleading signifier:* led confusion for the mappings (e.g., line coming out of their hands misassumed for selection) *Misaligned assumptions* (position, orientation, pressure): needed for placing correct gesture on virtual object (e.g., keyboard) due to lack of feedback
Insight Lungs	✅	*Visual affordance* (e.g., calibration screen showing tap + hold) helped get started	❌	*Ambiguous feedback* (e.g, while calibrating virtual objects): did not work as expected leading users to struggle with AR features/ interactions. *Misleading signifier* (e.g., AR feedback of pointing to floor): impacting the user perception of mappings b/w input and results
IKEA Place	✅	*Visual affordance* (e.g., calibration screen showing tap + hold) helped get started	❌	*Ambiguous or inadequate feedback:* led users to struggle understanding AR features (e.g., calibration ring) *Misleading signifier:* even after following AR feedback of pointing to vacant space, etc, participants were unclear why these errors occurred.

Fig. 9. Instances of how AR applications follow/ do not follow the NN design principles for gestural interfaces (Part 1). Both of the immersive AR apps violated the NN principles for the discoverability of gestural interactions and, as a result, our participants experienced high rates of both system and human errors. In contrast, users had an easier time getting started with both mobile applications as they incorporated visibility of system status. Users only experienced calibration and recognition issues with both mobile applications due to the misleading signifiers. (*Green tick mark implies followed NN principle; Yellow symbol implies partially followed NN principle; Red cross implies violated NN principle.)

	NN Principle: Reliability		NN Principle: Consistency and Standards	
Application Name	**Y/N**	**Evidence showing interface followed or violated NN principle**	**Y/N**	**Evidence showing interface followed or violated NN principle**
Graffiti 3D	❌	*Accidental activation:* wrong gesture (e.g., head/hand movement) mapped to correct outcome (e.g., invoking a menu) *Lack of consistency in the outcomes from the same gestures* (e.g., palm gesture for getting menu): led users to backtrack	❌	Input gestures lacked *consistency and transfer with real-world tasks/with 3D objects*: led to misassumptions on wrong gestures expected to produce a correct outcome (e.g., fist for grabbing, assumed interactions to be similar with physical/3D mechanical interfaces in terms of shape, depth perception)
Holomeeting	❌	*Misleading signifiers* (e.g., cursor line): leading to inaccurate explanations of how AR application and interactions works. *Random results of activation and lack of consistency* in the outcomes led to misconceptions on mappings (head/hand movements to move or access) virtual objects	❌	*Negative transfer b/w holomeeting tasks:* Similar task-specific gestural interaction (e.g., grabbing, rotating) offered less transfer b/w tasks to replicate a gesture for similar tasks.
Insight Lungs	⚠️	*Familiarity*: easier to get started with touch -based input *Calibration and recognition issues:* due to lack of feedback and the application's failure to recognize real-world objects	⚠️	*Positive transfer from mobile-based gestures*, but still users struggled because unlike non-AR apps everything in AR was moving in the space (e.g., users discovered swipe and zoom in gesture, but struggled to use them to interact with 3D objects, such as, 3D anatomy in AR).
IKEA Place	⚠️	*Familiarity*: easier to get started with touch -based input	⚠️	*Positive transfer from mobile-based gestures*, but still users struggled because

Fig. 10. [Continued] Instances of how AR applications follow/ do not follow the NN design principles for gestural interfaces (Part 2). Users experienced high rates of system and user-led errors with immersive AR apps as the apps lacked consistency and users experienced mostly negative transfer from real-world interactions. Mobile AR apps were more consistent with existing non-AR apps enabling positive transfer. Still, users experienced calibration and recognition issues due to the violation of the reliability principle. (*Green tick mark implies followed NN principle; Yellow symbol implies partially followed NN principle; Red cross implies violated NN principle.)

6.2 Transfer of Learning Between Tasks and Applications

Although the order of tasks was randomized for each participant, we observed some cases of positive and negative learning transfer. About half of the participants (10/25) tried to replicate and apply gestures from a previous task. Most of these positive transfer instances occurred during the use of the Graffiti 3D application, where participants tried to transfer the pinch gesture they learned during the sketching task to the grabbing and rotation tasks. In some cases, there was positive transfer across applications as well. For example, some participants applied similar gestures from the selection and menu invocation tasks from Graffiti 3D to Holomeeting: *"I will just do as I used in the sketch pen in the graffiti tool to see if there is a 3D pop up menu"* (P8).

Despite some positive transfer, several participants (14/25) experienced negative transfer. In most cases, these participants did not know where to replicate a gesture or how to apply their knowledge to similar tasks. For example, P3 faced negative transfer between the grabbing and rotation tasks within Holomeeting, e.g., *"maybe I need to select it, just as I did with sticky note... I cannot interact with it as easy as I thought, maybe it is two hands [for grabbing], because this method worked for sticky note, select it and grab it."* (P3).

Although there was evidence that positive transfer could help people learn unfamiliar gestures in AR, negative transfer could also hinder the learning process.

6.3 Issues Seeking Help

During the immersive-based AR tasks, almost everyone (23/25) sought help online (e.g., on Google, YouTube) but only one participant was successful at finding help and applying it to their tasks. In contrast, since mobile-based AR was more intuitive for participants, only 8 participants sought out help online (for a total of 13 help-seeking attempts), mostly for calibration (e.g., calibrating virtual object with respect to real-world surroundings) and advanced task interactions.

There were a total of 56 help-seeking attempts (Graffiti 3D: 21 attempts; Holomeeting: 35 attempts). Unlike the feature-specific, help-seeking attempts observed with 2D and 3D feature-rich applications in prior research [55, 66], almost half of the help-seeking attempts (24/56) in immersive AR were focused on understanding the interface itself and how to get started with gestural interactions, e.g., *"how to interact with Holomeeting app"* (P14), and *"getting started with graffiti 3D with Hololens"* (P5). The remaining help seeking attempts focused on other task-related queries about the basic gestures (25/56 attempts; 15/25 participants) or the advanced gestures for participants who reached that point (7/56 attempts; 6/25 participants). For example, participants formulated queries about learning gestures for selection, getting a menu, and grabbing, e.g., *"how to click while using AR whiteboard"* (P17), and *"move ruler holomeeting"* (P10). In fact, 5/25 participants (range of attempts: 4-8) performed end-to-start troubleshooting following help.

Despite these help-seeking efforts, only 2/56 attempts were successful and the remaining attempts either offered negative transfer (5/56) or no transfer (49/56), e.g., *"I mean rather than using the hands, I can also use the eye tracking sign according to some documentation in Google."* (P7). We observed that 11/25 participants failed to locate help in immersive AR because users lacked the correct mental model (seen in Section 4 and 5.3) and struggled to use the appropriate terminologies while framing queries. They mostly resorted to additional trials to figure out the more precise position, orientation and angle of the correct gesture: *"I just Googled directly...I tried..all the instructions. And it [gesture] sort of slightly worked, [but] then I'm doing trial-and-error and trying to change gestures like trying to hold the sketch from different angles so that it can move."* (P14).

7 EXPERT VALIDATION AND IMPLICATIONS

To substantiate and refine the interpretations drawn from our observation study, we conducted an expert validation [84] with 8 professional designers, engineers, and researchers in the AR domain.

7.1 Study Procedure and Participants

We recruited 8 experts (5M|3F) from four different large technology companies, including productivity, education, and gaming. All have dedicated product and research teams focused on AR devices and applications. Our participants had a variety of experiences with AR: designing AR applications and/or gestural interactions (>5 years on average), user experience research with AR (>5 years on average), and developing AR interfaces (2-5 years on average). All participants had graduate-level education in subjects, including CS (3/8), design (2/8), engineering(1/8), and information Science(1/8). They had worked on both immersive AR technologies (7/8), such as Hololens 2, Meta Quest Pro, and Google Glasses, as well as mobile AR technologies (5/8), such as, AR kit with iOS on an iPad or a phone. Furthermore, these experts had experience working on different aspects of AR, including interaction design and UX research (4/8), hardware development of devices (2/8), and input modeling using computer vision and digital signal processing (2/8).

We conducted semi-structured interviews lasting 45 minutes on average. We first shared our key observational findings through a slide presentation and asked each interviewee to reflect on these findings based on their own professional and research experience. We asked them to prioritize the importance of the issues for better supporting novice users of AR. We transcribed, coded, and analyzed the interview data using an inductive analysis approach [26].

7.2 Expert Reflections and Design Recommendations for Discoverable AR Gestural Interfaces

All of the experts agreed that the discoverability issues we uncovered involving system errors and human errors were commonplace with gestural interactions among novices. Although the experts had intuition about many of these errors and issues, they expressed that our study results helped substantiate these issues: *"[we] all know [these problems] are there but have not really scientifically categorized [them]"*(D08). Most participants mentioned discoverability in AR has unique challenges since the interaction language tends to be more "embodied and natural." The experts had anticipated that mobile AR would have a better learning transfer than immersive AR owing to "legacy bias (D05)" and "cultural propagation of these devices (D03)". Some of the experts also shared their experience with observing users in immersive AR, pointing out how users often gravitate towards their pre-existing mental models from desktop or web computing (D02). Unlike 2D/3D interfaces, users need to deal with the *"depth environment"* (D03) and *"excess degrees of freedom... comparing how people click in a mouse vs. use AR"* (D07).

We synthesized 3 key implications of our findings and identify opportunities for AR research and practice to addresses issues related to the interaction design, system recognition, and onboarding, summarized in Figure 11. These insights further corroborate the need to design AR interactions that better follow fundamental HCI design principles, summarized in Figures 9 and 10.

7.3 Addressing Issues at the Design Stage with Better Understanding of Mental Models

Our observational study showed how novices struggled at the outset due to their lack of familiarity with immersive mid-air gestural interactions, the lack of visibility within the AR applications, and their expectations from prior 2D/3D interactions.

The majority of AR experts (6/8) highlighted their own struggles when onboarding with AR and emphasized the significance of helping users form correct mental models. They expressed that the types of true negatives observed in our study provides evidence for how incorrect mental models impacted novices' understanding of the AR system at the basic level and resulted in misassumptions about "what they should do to fix the problem". As one AR expert explained, *"the mental model matters... people make workarounds that are not appropriate, such as pressing harder, when they don't* *understand what the system is actually looking for..."* (D03). The experts were not surprised to see novices struggle and highlighted the need to making gestural interaction in AR more accommodating of diverse user mental models: "we should be brainstorming as a community to try and think about how can we make sure that we're making technology that is accommodating the wide variety of people that are going to be using it" (D07).

All of the experts suggested that issues related to user errors need to be addressed at the design stage and provide feedback to guide users toward correct interactions when they are going down the wrong path. For example, D05 suggested one possible design direction by using *feedforward* to aid discoverability [30, 47, 48, 73, 76, 106] of different UI controls and corresponding actions. Unlike mobile AR applications, immersive AR apps lack consistency of gestures across application systems which makes learning harder. As users assumed to transfer their knowledge from real-world to the complex gesture sets, some experts suggested "simplified and standardized gesture sets (D07)" elicited via guessability studies [101] with novice users that can make the interactions simpler and easier to guess. Although prior research has attempted to design user-defined gesture sets [110] from trained users' perspectives, our results provide initial evidence that they do not translate the same way to novice users who form various misconceptions. To mitigate the misassumptions due to confusing mappings, 5/8 experts suggested the need to incorporate intuitive affordances and visibility of the system in the form of *"multimodality feedback (e.g., audio, visual or haptics)" (D06)* [31, 57]. Furthermore, another opportunity is to improve the design of interaction guidance, affordances (e.g., ghost hands) and mappings, and system feedback in the form of proper instructions, visual labels and multimodal feedback, or even basic visible UI (menu/screen) to minimize users' gulf of evaluation and gulf of execution. In particular, this will help participants who were somewhat close to performing the correct gesture (e.g., in Selection: 26% across 10 participants were tapping with one hand/both hands) but were unclear about how to correctly execute the gesture.

Some experts (3/8) highlighted that some errors are perhaps unavoidable in unfamiliar interfaces. But, they were optimistic about understanding and incorporating the user's mental model of the real world in these immersive applications: *"...the ultimate goal would be that people can just be natural about the way they're working with this [immersive application]"* (D06).

7.4 Addressing Issues at the Onboarding Stage and Improving In-Context Help

The novices in our study preferred to immediately explore interactions with AR on their own and neglected to seek any help or watch tutorials, eventually getting caught up in lengthy episodes of trial-and-error consistent with software learnability studies [22, 55, 80]). The majority of experts (7/8) were not surprised with this observation: *"I think we have to force a tutorial, honestly...people have this strong tendency to not spend time on initial tutorials or initial help"* (D08). The majority of the experts (7/8) felt that long tutorials in the beginning are not effective: *"they often blow up the application as they are long"*(D05), *people turn them [tutorial] off... they get boring." (D06). Tutorials also require users to "memorize a lot of information that...they do not know how they are going use until they are into the experience" (D06)* and users have no idea on *"how to trigger that tutorial on demand"*(D04). Having said that, they highlighted that some sort of "natural way" of onboarding is required when dealing with unfamiliar interactions in AR.

Many experts (5/8) highlighted that the onboarding process needs to be step-by-step, in-context and interactive analogous to the onboarding in video games: *"We can borrow a lot from what video games do in terms of walking people through the simple steps of doing something."*(D03) There is an opportunity in considering a staged approach for onboarding and designing gamifying walkthrough that can teach users gestural interactions through stages. Providing on-demand automated help that supports interactive guidance for mental models could also be worth exploring: *"offer*

Key Takeaways Observational Study and Interviews	Expert Validation and Implications
• *Difficulty at the outset in constructing an accurate mental model* of interacting with immersive AR • Difficulty in completing the study tasks as users could not accurately perform the necessary gestural interactions	• Incorrect mental models can diminish the uptake of products • More investment is needed in *empirically understanding AR consumers' mental models*, especially when they come from *diverse educational and professional backgrounds*
• Onboarding issues with gestural interactions could be attributed to a *gulf of execution* (e.g., gaps in figuring out the correct sequence of actions) or *gulf of evaluation* (e.g., gaps in evaluating the effect of an action), often due to lack of appropriate visual guidance and system feedback	• More effort needed at the design stage to make gestural interactions intuitive using techniques such as, *feedforward, simplified and standardized gesture sets, affordances and mappings, and multimodality feedback, interactive guidance*
• The majority of errors during gestural interaction across all tasks were were due to *human errors (true negatives)* and some due to *system-related errors (false positives and negatives)* • Users made incorrect assumptions and formed misconceptions based on their perceptions of other systems and real-world objects and due to confusing mappings and incorrect execution of gestures	• Many system-level errors can be resolved by developing gesture recognizers with high position accuracy, or train recognizers to detect error cases (error recognizers) • True negatives present an opportunity to *rethink true negatives as part of the interaction process and how to use their occurrence* to guide users toward correct interactions
• Users consistently preferred to *explore the interface on their own* rather than follow instructions or tutorials at the outset • When errors occurred with gestures, the majority of participants sought help by searching online, but most help-seeking *attempts were not successful* due to problems with formulating queries and applying the relevant instructions	• Errors are sometimes unavoidable in unfamiliar interfaces and users can try unexpected interactions. • It is important to consider appropriate *in-context and interactive onboarding* process (analogous to the onboarding in video games) and *training* that, unlike tutorials, focuses on training users step-by-step in a natural way.

Fig. 11. Summary of Key Takeaways and Implications. *Each row represents key takeaway (on the left) and corresponding implication (on the right)

help on the fly...maybe something subtle like a little icon appears in the corner that says like you know here's an animation of the gesture you're supposed to do."(D07)

Some experts (2/8) advocated for onboarding using metaphors and design gestures that facilitate natural mappings: *"...drawing on like physical reality and metaphor to design the gestures in the first place will aid in the discoverability and memorability of gestures"*(D07). Other domains such as 2D and 3D applications have already explored the design of in-context help [19, 25, 43, 46, 105]; however, there has been little research into designing 3D gestural guides in AR that provide the visual gestural feedback observed in 2D gestural interactions [10, 13, 18, 38] (e.g., mouse-based interactions, table-top hand gestures). Given the complexity and unique interaction challenges with mid-air interactions for novices identified in our study, future work can look at expanding the scope of in-context help solutions that can facilitate AR-specific onboarding.

7.5 Addressing Issues at the System Level and Improving Machine Recognition

The analysis of participants' discoverability attempts and errors has several implications to improve system-level machine recognition and create the next-generation of understandable AR user interfaces. Given the impact of false negatives on trial and error behaviors, one obvious outcome suggested by 6/8 experts is the need for *better gestural recognizers* so that there is high position accuracy and precision in recognizing that the gesture did not produce any outcome. The experts asserted that although recognizers can never be perfect, it is important to properly train gesture recognizers [96] to be more adaptive and representative of real-world datasets that simulating with more samples from novices. This type of data can mitigate false positives and their impact on the learnability of gestures: *"...if pinch recognition is not working, perhaps you may design your classifier to be adaptive that makes the model more reliable or robust for that context."*(D07) Another opportunity is to train recognizers to detect error cases [3] *(error recognizers)*, or identify when users get into trial and error loops. Some (3/8) experts further argued that incorporating error data from nuances experienced by novices has the potential to build system understanding for detecting those errors: *"I would advocate for... let's detect when things are going wrong. Kind of bespoke guidance about do not do that do this instead"* (D06). These error recognizers can further be personalized by modelling them with data from individual users: *"I think we need to build those cycles in and then basically the system needs to learn...what kinds of errors are systematic to a particular user?"*(D04)

We observed clusters of participants performing similar false positive gestures (Figure 5) for certain tasks, suggesting that there may be patterns of negative learning transfer bolstered by false positives that eventually lead to incorrect interactions. These findings can be useful when preparing error datasets and modeling incorrect interactions to facilitate user learning or error recovery. Given the black-box nature of the sensing technologies to the users as well as designers, some experts (2/8) suggested that providing training to users about the underlying sensing technologies can help in understanding the system better and aid discoverability.

8 LIMITATIONS

The findings of this research were based on an observational study and a qualitative analysis of the resulting research data. To limit researcher bias, we ensured inter-observer reliability in the coding. Although we sampled from a diverse range of novice AR users from different educational and professional backgrounds, they were all living in North America at the time of the study and had verbal fluency in English. The insights were drawn from a gestural error analysis that utilized data from four broad tasks that used unimanual and bimanual gestural interactions. Future research should expand upon these insights and examine other possible tasks that could be used to understand user behavior and even train gesture recognizers. While the expert validation showed that the discoverability issues uncovered in our study were "real problems" based on their experiences with AR design and implementation, future studies should experimentally validate these findings on other commercial AR devices, applications, tasks and interactions that were not considered in the current study. Our findings synthesize key barriers of learnability over a single snapshot of time, thus future research should apply more longitudinal approaches to quantify the learning curve with these AR applications over a period of time. Future research can also compare the discoverability issues that users who have prior experience with AR, face in these environments. Although the AR experts in our validation study confirmed that users tend to skip instructions and long tutorials, in future work it would be interesting to investigate whether in-context and interactive onboarding instead could lower the barriers to properly learning AR gestural interactions.

9 CONCLUSIONS

The rapid advancement of AR has sparked substantial interest in both industry and research to drive widespread consumer-level adoption by focusing on issues related to usability and user experience [29, 41, 54, 81–83]. Our in-lab observational study complements these ongoing efforts by contributing empirical insights into how 25 novice everyday consumers with varied technical skills form (often incorrect) mental models of immersive AR applications and struggle with discovering the correct gestural interaction. Our results highlighted the various types of errors (i.e., false positives, false negatives, and true negatives) that users experienced during gestural interactions and how these errors impeded users' ability to complete a given task. Even though users experienced false negatives with their continued trial-and-error cycles, we found that false positives catastrophically impacted the discoverability of gestural interaction, similar to research on mouse-based interaction [62]. Our study highlights the importance of understanding users' gestural interactions and misconceptions and how they form their mental model on encountering errors, which has long been advocated as the first step in facilitating interaction with unfamiliar modalities [75]. As seen in prior works [22, 55, 80], we also observed that novices were eager to explore the interface on their own rather than follow instructions or tutorials, which has important implications for the design of more intuitive gestural input and interaction and more streamlined in-context help and onboarding opportunities.

Overall, our work has contributed a detailed picture of how a diverse group of users began working within unfamiliar AR environments, discovered gestural-based input and interactions, encountered breakdowns in discoverability, and struggled to locate and apply relevant help. We validated the observational findings with 8 professional AR designers, engineers, and researchers and highlight opportunities for engineers or designers working on devising new gestural interactions in AR, UX practitioners in AR, machine learning engineers improving gestural recognizers and other researchers enhancing the user experience and adoption of AR applications among diverse novice users.

10 LIST OF REVISIONS IMWUT 23:

We thank the reviewers for their thoughtful comments and for appreciating our work. We offer a few clarifications and list the revisions that we have made based on the valuable reviewer suggestions.

- **(1AE, 2AE, R3, R4) All reviewers asked for specific implications/lessons for AR interaction designers from our findings. 1AE suggested to include a "second study" to create specific suggestions:** We agree with the reviewers' concerns and appreciate suggestions by 1AE to include an additional study. So, to validate our observations and interpretations and better understand the specific implications for practice, we did an expert validation study with 8 professional AR designers, engineers, and researchers from four different large technology companies that build AR-related software technologies. We presented to them a summary of our key findings and asked them to reflect, critique, and share their experiences for improving gestural interaction with AR. Although experts had intuition about many of the errors and discoverability issues that we uncovered, they expressed that our study results helped substantiate that these issues *"are real problems."*.
  (1) Based on the feedback provided by our experts and our own interpretation, we synthesize 3 key implications of our findings and identify more concrete and actionable opportunities for AR designers, engineers, researchers, and the broader HCI community to create the next-generation of input for AR user interfaces in three directions: (i) Addressing Issues at the Design Stage with Better understanding of Mental Models (ii) Addressing Issues at the Onboarding Stage and Improving In-Context Help (iii) Addressing Issues at the System Level and Improving Machine Recognition. We reframed Section 7 to be "Expert Validation and Implications" (For the detailed results and implications, See Section 7).
  (2) Furthermore, the insights obtained from the experts also validate the need to design AR interactions that better follow fundamental HCI design principles (discussed in Figure 9, 10).
- **(1AE, 2AE, R4) The reviewers asked to discuss if and how order effects were handled in our study:** We clarified our definition of random ordering in Section 3.4.2, we did in fact counterbalance the order to reduce order effects. Each participant completed two tasks (with basic and advanced interactions) for each application on two different devices (immersive and mobile) in a random order. The order in which tasks were presented to

each study participant was determined using a Latin Square where 8 orders were possible for assignment. (See Section 3.4.2).

- **(2AE, R3) 2AE and R3 wished for more description of the differences the authors observed when it comes to participants' prior knowledge (expert tech users) and the feedback the different groups provided:**
  (1) We apologize for the confusion regarding our use of the term "expert users." These 4 participants we recruited in our study did not have any experience with AR environments, so we removed the reference of expert users and updated Section 3.1, as follows, we recruited 4 additional participants who did not have experience with AR, but had prior experience with other forms of 3D interaction (e.g., 3D design or playing VR games). (See Section 3.1).
  (2) Having said that, we understand the reviewers feedback on including description of differences observed when it comes to participants' prior knowledge with VR/3D space. We highlighted the results from these more experienced participants and more clearly described the context around their quotes in Section 5. We observed that these users got started with the interface accidentally or following the affordances from the AR application. But eventually, even after having prior experience in interacting with VR games or 3D interfaces, these participants struggled the same way as other users with misleading signifiers, confusing mappings and trial-and-error loops of system and user errors. These results underscore the struggle users experienced in AR environments, which is different from other VR or 3D interfaces and emphasize that the prior knowledge of VR did not translate the same way to AR environments. (See Section 5 and 5.2)
  (3) We included in the section 8 that future research can also compare the discoverability issues that users who have prior experience with AR, face in these environments (See Section 8).
- **(1AE) 1AE suggested to restructure the paper discussing N&N's fundamental design principles, how many gesture interfaces do not follow them along with ranking interactions based on these design principles:** We appreciate and followed this suggestion to better present our results, as follows,
  (1) We included Figure 9 and 10 to represent findings on what and how these gesture interfaces follow/do not follow necessary N&N's fundamental design principles. Furthermore, in both these figures, we have ranked the interactions studied by how well they follow the principles using the visual coding symbols, i.e., Yellow symbol implies partially followed NN principle ; Green tick mark implies followed NN principle; Red cross implies violated NN principle. (See Figure 9, Figure 10)
  (2) In addition, we also included information in Section 5 about how many interfaces did not follow these principles and how violating these principles, as seen in immersive AR apps, led to high rates of both system and human error impacting the discoverability. In contrast, mobile AR applications had followed most of these design principles (e.g., visibility of system status and consistency with existing non-AR apps) allowing users to have easier time getting started with them (See Section 5).
  (3) Furthermore, these findings on the need to design interactions that better follow fundamental HCI design were validated by the experts during our second study in Section 7.2. (See Section 7.2)
  (4) To ensure consistency and clarity throughout the paper, we introduced the NN design principle for gestural interfaces in the Section 2.2, Section 3.5.
- **(R3) R3 asked for additional description on how people held/interacted with the mobile devices in terms of set up for the mobile devices, device size and weight:** We clarified the device size and weight

in Section 3.2: Google Pixel 6 Android phone (6.2" X 2.9" X 0.4" with a 1080 x 2400 OLED resolution and 207 grams weight) and an iPad mini (7.69" X 5.3" X 0.25" and 297 grams weight with a 2266 X 1488 resolution).

(1) In terms of the device setup, we clarified that for both mobile and AR applications, the users were handed the devices with the application opened by the researcher so that they could focus on the given tasks. (See Section 3.4.2)

(2) Additionally, we included a Figure 4 explaining how users held/interacted with mobile devices i.e., description of gestures when performing given tasks in mobile AR applications. (See Figure 4)

- **(1AE, R3) had concerns if setup/onboarding could help familiarize users with AR:** We agree with the reviewers' concern and therefore, we asked our experts in the second study to reflect upon our finding on participants not following tutorials and jumping into the AR interfaces.

  (1) We clarified in Section 7.4, wherein we found that, although the majority of experts (7/8) agreed with having more onboarding options, they were not surprised that our participants did not seek tutorials or help. In fact, the experts confirmed that users tend to skip long tutorials even when it is beneficial for them. In the implications, we do discuss ideas for in-context and interactive onboarding instead of long tutorials to get started with the AR gestural interactions that the experts suggested could yield more user participation. (See Section 7.4)

  (2) We also clarified in Section 3.4.1 why we did not force participants to watch setup tutorias, as we focused on more sophisticated range of task-specific gestural interactions and were not offered in the setup tutorial. Having said that, users were given freedom to seek help and watch more in-context tutorials or help from online resources (See Section 3.4.1).

- **(R4, 1AE) R4 and 1AE suggested us to address why do (or do not) the results on the Hololens generalize to other headsets:** This is a good point and based on our findings from Expert validation study, we discussed this in our revisions in Section 8. Our experts who had worked with a wide range of commercial AR devices (e.g., Hololens, Meta Quest Pro, Google Glasses) could relate to the discoverability issues uncovered in our study and confirm them to "real problems" across these platforms. In fact, they also suggested these findings could be generalizable to immersive AR. We highlighted that the future studies may, however, experimentally validate these findings on the other commercial AR devices, applications, tasks and interactions that were not considered in the current study. (See Section 8)
- **(R4, 1AE) R4 and 1AE suggested us to better motivate why the 4 apps were chosen:** We clarified and provided more information on our criteria followed to chose these applications for the user study in Section 3.3, as follows, To better motivate diverse set of novice users to perform tasks with AR applications during the study, we focused on popular productivity-related or day-to-day tasks that everyday consumers could find relevant and we avoided niche gaming applications. We explored applications that were already popular on the Hololens 2 platform [92] for boosting productivity or performing day-to-day tasks related to education, remote collaboration, shopping, among others.(See Section 3.3)
- **(R4, 1AE) R4 and 1AE asked for more information on how participants were introduced to AR:** In Section 3.4.1, we included more details about how users were handed the device and application during the study, as follows:For both mobile and immersive AR applications, the users were handed the devices with the application opened by the researcher so they could focus on the given tasks. In the case of immersive AR, the researcher helped the user to wear the head-mounted display with the application open (See Section 3.4.1).

- (R3) **R3 suggested to include a summary table of all the challenges faced, mitigations, and recommendations to summarize the findings:** We included Figure 11: Summary of Key Takeaways and Implications showing the challenges faced by our participants and mitigations/recommendation validated by our experts (See Figure 11).
- **(R3) R3 suggested to discuss any differences and expectations between system-level onboarding vs. expecting the apps to do onboarding or providing "in-context" instructions:**Our second study with experts provided detailed insights on the difference and expectations *in practice* for onboarding at the system-level or in-context, which we discussed in Section 7.4. The majority of experts (7/8) did not favour for setup tutorials or system-level onboarding and expressed that long setup tutorials of onboarding that comes in the beginning as " often blow up the application as they are long"(D05) and users turn them off: " people turn them [tutorial] off... they get boring and [users] would not memorize a lot of information that might not be broken or they do not know how they are gonna use it until they are into the experience" (D06) and users have no idea on "how to trigger that tutorial on demand"(D04). Instead, they expressed that some sort of natural way of onboarding is required when dealing with unfamiliar interactions such as AR which users are not used to. This was also in consistent with (6/25) participants who suggested that it would be helpful to have proper in-context onboarding activities and instructions from the start as a way of initially becoming familiar with AR and gestural input (that go beyond setting up the hardware device). (See Section 7.4)
- **((R4) R4 asked for clearly defining terms such as immersive AR and mid-air gestures in the Introduction:**
  (1) We clarified the definition of immersive AR applications in Section 1 as well as Section 3.2, as applications hosted on head-mounted displays (e.g., Microsoft Hololens, Magic Leap). Furthermore, we added: "the complexity intensifies in the domain of immersive AR hosted on head-mounted displays (e.g., Microsoft Hololens, Magic Leap), where sensors and cameras are used to seamlessly integrate 3D virtual objects into the user's immediate physical surroundings, blurring the boundaries between the digital and physical realms." (See Section 1 and Section 3.2)
  (2) We also clarified mid-air gesture definition as: mid-air gestures, where users can interact with the user interface without touching or holding a physical device. (See Section 1)

10.1 Formatting changes

We appreciate all the formatting changes and typos pointed out by the reviewers. We highlight some of these changes here (especially those explicitly outlined in the reviews) as well as major changes to Section 7.

- **(1AE R4) R4 and 1AE asked to tighten the paper and do proofing:** We performed a full scan to tighten the paper and did proofing. As suggested by R4, we tighten the section 6.3 and made it more concise.
- **(R3)** Fixed missing figure references (with ?) in the Section 5.1 and fixed minor typos/grammar issues in a full pass of the paper.

REFERENCES

[1] 2019. Get hololens tips - microsoft store en-CA. https://www.microsoft.com/en-ca/p/hololens-tips/9pd4cxkklc47?activetab=pivot%3Aoverviewtab

[2] 2020. Get holomeeting 2. https://www.microsoft.com/en-us/p/holomeeting-2/9nh5zdb4ggbq?activetab=pivot%3Aoverviewtab

[3] 2021. Responsible machine learning with error analysis. https://techcommunity.microsoft.com/t5/ai-machine-learning-blog/responsible-machine-learning-with-error-analysis/ba-p/2141774

[4] 2022. Insight Lung - apps on Google Play. https://play.google.com/store/apps/details?id=com.animares.insightlung&hl=en&gl=US

[5] Thomas Alsop. 2022. Augmented reality (AR) - statistics amp; facts. https://www.statista.com/topics/3286/augmented-reality-ar/#topicOverview

[6] Christopher R. Austin, Barrett Ens, Kadek Ananta Satriadi, and Bernhard Jenny. 2020. Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality. *Cartography and Geographic Information Science* 47, 3 (2020), 214–228. https://doi.org/10.1080/15230406.2019.1696232 arXiv:https://doi.org/10.1080/15230406.2019.1696232

[7] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. 2001. Recent advances in augmented reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47. https://doi.org/10.1109/38.963459

[8] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (08 1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355 arXiv:https://direct.mit.edu/pvar/article-pdf/6/4/355/1623026/pres.1997.6.4.355.pdf

[9] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. 2012. ShoeSense: A New Perspective on Gestural Interaction and Wearable Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1239–1248. https://doi.org/10.1145/2207676.2208576

[10] Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (Monterey, CA, USA) *(UIST '08)*. Association for Computing Machinery, New York, NY, USA, 37–46. https://doi.org/10.1145/1449715.1449724

[11] Sana Behnam and Raluca Budiu. 2022. The usability of augmented reality. https://www.nngroup.com/articles/ar-ux-guidelines/

[12] Sana Behnam and Raluca Budiu. 2022. The usability of augmented reality. https://www.nngroup.com/articles/ar-ux-guidelines/

[13] Mike Bennett, Kevin McCarthy, Sile O'modhrain, and Barry Smyth. 2011. Simpleflow: enhancing gestural interaction with gesture prediction, abbreviation and autocompletion. In *IFIP Conference on Human-Computer Interaction*. Springer, 591–608.

[14] Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. 2014. Exploring Smartphone-Based Interaction with Overview+detail Interfaces on 3D Public Displays. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices amp; Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 125–134. https://doi.org/10.1145/2628363.2628374

[15] Zunaira Ilyas Bhutta, Syedda Umm-e Hani, and Iqra Tariq. 2015. The next problems to solve in augmented reality. In *2015 International Conference on Information and Communication Technologies (ICICT)*. 1–4. https://doi.org/10.1109/ICICT.2015.7469490

[16] Mark Billinghurst. 2021. Grand Challenges for Augmented Reality. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.578080

[17] Mark Billinghurst, Adrian Clark, Gun Lee, et al. 2015. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272.

[18] Andrew Bragdon, Robert Zeleznik, Brian Williamson, Timothy Miller, and Joseph J LaViola Jr. 2009. GestureBar: improving the approachability of gesture-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2269–2278.

[19] Joel Brandt, Mira Dontcheva, Marcos Weskamp, and Scott R. Klemmer. 2010. Example-Centric Programming: Integrating Web Search into the Development Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 513–522. https://doi.org/10.1145/1753326.1753402

[20] Eugenie Brasier, Olivier Chapuis, Nicolas Ferey, Jeanne Vezien, and Caroline Appert. 2020. ARPads: Mid-air Indirect Input for Augmented Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 332–343. https://doi.org/10.1109/ISMAR50242.2020.00060

[21] Wolfgang Buschel, Jian Chen, Raimund Dachselt, Steven Drucker, Tim Dwyer, Carsten Gorg, Tobias Isenberg, Andreas Kern, Chris North, and Wolfgang Stuerzlinger. 2018. *Interaction for Immersive Analytics*. Vol. 11190. SpringerLink, 95–138. https://www.microsoft.com/en-us/research/publication/interaction-for-immersive-analytics/

[22] John M. Carroll and Mary Beth Rosson. 1987. *Paradox of the Active User*. MIT Press, Cambridge, MA, USA, 80–111.

[23] Edwige Chauvergne, Martin Hachet, and Arnaud Prouzeau. 2023. User Onboarding in Virtual Reality: An Investigation of Current Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 711, 15 pages. https://doi.org/10.1145/3544548.3581211

[24] Yuan Chen, Keiko Katsuragawa, and Edward Lank. 2020. Understanding Viewport- and World-Based Pointing with Everyday Smart Devices in Immersive Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376592

[25] Parmit K. Chilana, Amy J. Ko, and Jacob O. Wobbrock. 2012. LemonAid: Selection-Based Crowdsourced Contextual Help for Web Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1549–1558. https://doi.org/10.1145/2207676.2208620

[26] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1 (1990), 3–21.

[27] Meghal Dani, Gaurav Garg, Ramakrishna Perla, and Ramya Hebbalaguppe. 2018. Mid-Air Fingertip-Based User Interaction in Mixed Reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 174–178. https://doi.org/10.1109/ISMAR-Adjunct.2018.00061

[28] Arindam Dey, Mark Billinghurst, Robert W Lindeman, and J Edward Swan. 2018. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018), 37.

[29] Amandeep Dhir and Mohammed Al-Kahtani. 2013. A case study on user experience (UX) evaluation of mobile augmented reality prototypes. *Journal of Universal Computer Science* 19, 8 (2013), 1175–1196.

[30] Tom Djajadiningrat, Kees Overbeeke, and Stephan Wensveen. 2002. But How, Donald, Tell Us How? On the Creation of Meaning in Interaction Design through Feedforward and Inherent Feedback. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (London, England) *(DIS '02)*. Association for Computing Machinery, New York, NY, USA, 285–291. https://doi.org/10.1145/778712.778752

[31] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI principles to AR systems design. (2007).

[32] Rob Enderle. 2021. How mixed reality (and hololens) boosted productivity at Lockheed Martin. https://www.computerworld.com/article/3604508/how-mixed-reality-and-hololens-boosted-productivity-at-lockheed-martin.html

[33] Tristan Endsley, Kelly Sprehn, Ryan Brill, Kimberly Ryan, Emily Vincent, and James Martin. 2017. Augmented Reality Design Heuristics: Designing for Dynamic Interactions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61 (09 2017), 2100–2104. https://doi.org/10.1177/1541931213602007

[34] Margarita Esau, Veronika Krauß, Dennis Lawo, and Gunnar Stevens. 2022. Losing Its Touch: Understanding User Perception of Multimodal Interaction and Smart Assistance. In *Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1288–1299. https://doi.org/10.1145/3532106.3533455

[35] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27, 8 (2006), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010 ROC Analysis in Pattern Recognition.

[36] Michela Ferron, Nadia Mana, and Ornella Mich. 2019. *Designing Mid-Air Gesture Interaction with Mobile Devices for Older Adults*. Springer International Publishing, Cham, 81–100. https://doi.org/10.1007/978-3-030-06076-3_6

[37] Adam Fourney, Ben Lafreniere, Parmit Chilana, and Michael Terry. 2014. InterTwine: Creating Interapplication Information Scent to Support Coordinated Use of Software. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 429–438. https://doi.org/10.1145/2642918.2647420

[38] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. 2009. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM international conference on interactive tabletops and surfaces*. 165–172.

[39] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The Vocabulary Problem in Human-System Communication. *Commun. ACM* 30, 11 (nov 1987), 964–971. https://doi.org/10.1145/32206.32212

[40] Sharon Nelson-Le Gall. 1985. Chapter 2: Help-seeking behavior in learning. *Review of research in education* 12, 1 (1985), 55–90.

[41] Yahya Ghazwani and Shamus Smith. 2020. Interaction in Augmented Reality: Challenges to Enhance User Experience. In *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations* (Sydney, NSW, Australia) *(ICVARS 2020)*. Association for Computing Machinery, New York, NY, USA, 39–44. https://doi.org/10.1145/3385378.3385384

[42] Ari Grobman. 2018. AR's success depends on perfecting input methods. https://thenextweb.com/news/ars-success-depends-on-perfecting-input-methods

[43] Tovi Grossman and George Fitzmaurice. 2010. ToolClips: An Investigation of Contextual Video Assistance for Functionality Understanding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1515–1524. https://doi.org/10.1145/1753326.1753552

[44] Tovi Grossman, George Fitzmaurice, and Ramtin Attar. 2009. A Survey of Software Learnability: Metrics, Methodologies and Guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 649–658. https://doi.org/10.1145/1518701.1518803

[45] Aurora Harley. 2020. Augmented reality for ecommerce: Is it useful yet? https://www.nngroup.com/articles/augmented-reality-useful/

[46] Björn Hartmann, Daniel MacDougall, Joel Brandt, and Scott R. Klemmer. 2010. What Would Other Programmers Do: Suggesting Solutions to Error Messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1019–1028. https://doi.org/10.1145/1753326.1753478

[47] Rex Hartson and Pardha Pyla. 2019. Chapter 31 - The Interaction Cycle. In *The UX Book (Second Edition)* (second edition ed.), Rex Hartson and Pardha Pyla (Eds.). Morgan Kaufmann, Boston, 695–707. https://doi.org/10.1016/B978-0-12-805342-3.00031-X

[48] Rex Hartson and Pardha Pyla. 2019. Chapter 33 - Background: Affordances, the Interaction Cycle, and UX Design Guidelines. In *The UX Book (Second Edition)* (second edition ed.), Rex Hartson and Pardha Pyla (Eds.). Morgan Kaufmann, Boston, 825–832. https://doi.org/10.1016/B978-0-12-805342-3.00033-3

[49] Uta Hinrichs and Sheelagh Carpendale. 2011. Gestures in the Wild: Studying Multi-Touch Gesture Sequences on Interactive Tabletop Exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3023–3032. https://doi.org/10.1145/1978942.1979391

[50] Jason Hong. 2013. Considering Privacy Issues in the Context of Google Glass. *Commun. ACM* 56, 11 (nov 2013), 10–11. https://doi.org/10.1145/2524713.2524717

[51] John B. Horrigan and Sydney Jones. 2020. When technology fails. https://www.pewresearch.org/internet/2008/11/16/when-technology-fails/

[52] Nathaniel Hudson, Benjamin Lafreniere, Parmit K. Chilana, and Tovi Grossman. 2018. Investigating How Online Help and Learning Resources Support Children's Use of 3D Design Software. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173831

[53] niantic inc. 2022. Pokémon go - apps on google play. https://play.google.com/store/apps/details?id=com.nianticlabs.pokemongo&hl=en_CA&gl=US&pli=1

[54] Shafaq Irshad and Dayang Rohaya Bt Awang Rambli. 2014. User experience of mobile augmented reality: A review of studies. In *2014 3rd international conference on user science and engineering (i-USEr)*. IEEE, 125–130.

[55] Kimia Kiani, George Cui, Andrea Bunt, Joanna McGrenere, and Parmit K. Chilana. 2019. Beyond "One-Size-Fits-All": Understanding the Diversity in How Software Newcomers Discover and Make Use of Help Resources. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300570

[56] Marion Koelle, Abdallah El Ali, Vanessa Cobus, Wilko Heuten, and Susanne CJ Boll. 2017. All about Acceptability? Identifying Factors for the Adoption of Data Glasses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 295–300. https://doi.org/10.1145/3025453.3025749

[57] Panos E Kourouthanassis, Costas Boletsis, and George Lekakos. 2015. Demystifying the design of mobile augmented reality applications. *Multimedia Tools and Applications* 74, 3 (2015), 1045–1066.

[58] Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical Research in Mid-Air Interaction: A Systematic Review. *International Journal of Human–Computer Interaction* 35, 18 (2019), 1747–1768. https://doi.org/10.1080/10447318.2019.1572352 arXiv:https://doi.org/10.1080/10447318.2019.1572352

[59] Audrey Labrie and Jinghui Cheng. 2020. Adapting Usability Heuristics to the Context of Mobile Augmented Reality. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 4–6. https://doi.org/10.1145/3379350.3416167

[60] Benjamin Lafreniere, Parmit K. Chilana, Adam Fourney, and Michael A. Terry. 2015. These Aren't the Commands You're Looking For: Addressing False Feedforward in Feature-Rich Software. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software amp; Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 619–628. https://doi.org/10.1145/2807442.2807482

[61] Benjamin Lafreniere, Tovi Grossman, and George Fitzmaurice. 2013. Community Enhanced Tutorials: Improving Tutorials with Multiple Demonstrations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1779–1788. https://doi.org/10.1145/2470654.2466235

[62] Ben Lafreniere, Tanya R. Jonker, Stephanie Santosa, Mark Parent, Michael Glueck, Tovi Grossman, Hrvoje Benko, and Daniel Wigdor. 2021. False Positives vs. False Negatives: The Effects of Recovery Time and Cognitive Costs on Input Error Preference. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 54–68. https://doi.org/10.1145/3472749.3474735

[63] DoYoung Lee, Youryang Lee, Yonghwan Shin, and Ian Oakley. 2018. Designing Socially Acceptable Hand-to-Face Input. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 711–723. https://doi.org/10.1145/3242587.3242642

[64] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-Air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software amp; Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 63–71. https://doi.org/10.1145/2807442.2807489

[65] Google LLC. 2012. Google maps. https://apps.apple.com/us/app/google-maps/id585027354

[66] Shareen Mahmud, Jessalyn Alvina, Parmit K. Chilana, Andrea Bunt, and Joanna McGrenere. 2020. Learning Through Exploration: How Children, Adults, and Older Adults Interact with a New Feature-Rich Application. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376414

[67] Bernard Marr. 2021. The fascinating history and evolution of extended reality (XR) – covering AR, VR and mr. https://www.forbes.com/sites/bernardmarr/2021/05/17/the-fascinating-history-and-evolution-of-extended-reality-xr--covering-ar-vr-and-mr/?sh=76f8e854bfd6

[68] Pranav Mistry and Pattie Maes. 2009. SixthSense: A Wearable Gestural Interface. In *ACM SIGGRAPH ASIA 2009 Sketches* (Yokohama, Japan) *(SIGGRAPH ASIA '09)*. Association for Computing Machinery, New York, NY, USA, Article 11, 1 pages. https://doi.org/10.1145/1667146.1667160

[69] Denisse Moreno. 2020. Most Americans aren't sure what augmented reality is. https://www.ibtimes.com/facebook-apple-pushing-ar-many-americans-dont-know-what-it-2573065

[70] Denisse Moreno. 2020. Most Americans aren't sure what augmented reality is. https://www.ibtimes.com/facebook-apple-pushing-ar-many-americans-dont-know-what-it-2573065

[71] Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding Users' Preferences for Surface Gestures. In *Proceedings of Graphics Interface 2010* (Ottawa, Ontario, Canada) *(GI '10)*. Canadian Information Processing Society, CAN, 261–268.

[72] Florian Müller, Niloofar Dezfuli, Max Mühlhäuser, Martin Schmitz, and Mohammadreza Khalilbeigi. 2015. Palm-Based Interaction with Head-Mounted Displays. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 963–965. https://doi.org/10.1145/2786567.2794314

[73] Andreea Muresan, Jess McIntosh, and Kasper Hornbæk. 2023. Using Feedforward to Reveal Interaction Possibilities in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* (jun 2023). https://doi.org/10.1145/3603623 Just Accepted.

[74] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-Air Pan-and-Zoom on Wall-Sized Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 177–186. https://doi.org/10.1145/1978942.1978969

[75] Donald Norman. 1990. A.(1990) The Design of Everyday Things.

[76] Don Norman. 2013. *The design of everyday things: Revised and expanded edition.* Basic books.

[77] Donald A. Norman and Jakob Nielsen. 2010. Gestural Interfaces: A Step Backward in Usability. *Interactions* 17, 5 (sep 2010), 46–49. https://doi.org/10.1145/1836216.1836228

[78] David G. Novick, Oscar D. Andrade, and Nathaniel Bean. 2009. The Micro-Structure of Use of Help. In *Proceedings of the 27th ACM International Conference on Design of Communication* (Bloomington, Indiana, USA) *(SIGDOC '09)*. Association for Computing Machinery, New York, NY, USA, 97–104. https://doi.org/10.1145/1621995.1622014

[79] David G. Novick, Oscar D. Andrade, Nathaniel Bean, and Edith Elizalde. 2008. Help-Based Tutorials. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication* (Lisbon, Portugal) *(SIGDOC '08)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1456536.1456538

[80] David G. Novick and Karen Ward. 2006. Why Don't People Read the Manual?. In *Proceedings of the 24th Annual ACM International Conference on Design of Communication* (Myrtle Beach, SC, USA) *(SIGDOC '06)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/1166324.1166329

[81] Thomas Olsson, Tuula Kärkkäinen, Else Lagerstam, and Leena Ventä-Olkkonen. 2012. User Evaluation of Mobile Augmented Reality Scenarios. *J. Ambient Intell. Smart Environ.* 4, 1 (jan 2012), 29–47.

[82] Thomas Olsson, Else Lagerstam, Tuula Kärkkäinen, and Kaisa Väänänen-Vainio-Mattila. 2013. Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Personal and ubiquitous computing* 17, 2 (2013), 287–304.

[83] Thomas Olsson and Markus Salo. 2012. Narratives of Satisfying and Unsatisfying Experiences of Current Mobile Augmented Reality Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2779–2788. https://doi.org/10.1145/2207676.2208677

[84] Michael Q. Patton. 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research* 34 5 Pt 2 (1999), 1189–208. https://api.semanticscholar.org/CorpusID:15176061

[85] Martha Pease. 2015. Google Glass: What went wrong. https://www.cnn.com/2015/01/20/opinion/pease-google-glass-what-went-wrong/index.html

[86] Tran Pham, Jo Vermeulen, Anthony Tang, and Lindsay MacDonald Vermeulen. 2018. Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 227–240. https://doi.org/10.1145/3196709.3196719

[87] Krzysztof Pietroszek, Liudmila Tahai, James R. Wallace, and Edward Lank. 2017. Watchcasting: Freehand 3D interaction with off-the-shelf smartwatch. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 172–175. https://doi.org/10.1109/3DUI.2017.7893335

[88] Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark Billinghurst. 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 73–82. https://doi.org/10.1109/ISMAR.2014.6948411

[89] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 282–299.

[90] David M. W. Powers. 2008. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation.

[91] Michael L Raulin and Anthony M Graziano. 2019. Quasi-experiments and correlational studies. In *Companion Encyclopedia of Psychology*. Routledge, 1122–1141.

[92] Virtual Reality Experiences. 2022. 15 best hololens 2 apps. https://www.virtualrealityexp.co.uk/15-best-hololens-2-apps/

[93] Marc Rettig. 1991. Nobody Reads Documentation. *Commun. ACM* 34, 7 (jul 1991), 19–24. https://doi.org/10.1145/105783.105788

[94] John Rieman. 1996. A Field Study of Exploratory Learning Strategies. *ACM Trans. Comput.-Hum. Interact.* 3, 3 (sep 1996), 189–218. https://doi.org/10.1145/234526.234527

[95] Houssem Saidi, Marcos Serrano, Pourang Irani, Christophe Hurter, and Emmanuel Dubois. 2019. On-Body Tangible Interaction: Using the Body to Support Tangible Manipulations for Immersive Environments. In *Human-Computer Interaction – INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV* (Paphos, Cyprus). Springer-Verlag, Berlin, Heidelberg, 471–492. https://doi.org/10.1007/978-3-030-29390-1_26

[96] Julia Schwarz, Scott Hudson, Jennifer Mankoff, and Andrew D. Wilson. 2010. A Framework for Robust and Flexible Handling of Inputs with Uncertainty. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/1866029.1866039

[97] Scooley. 2021. Getting around hololens 2. https://learn.microsoft.com/en-us/hololens/hololens2-basic-usage

[98] Scooley. 2021. Set up your hololens 2. https://learn.microsoft.com/en-us/hololens/hololens2-start

[99] Shaishav Siddhpuria, Keiko Katsuragawa, James R Wallace, and Edward Lank. 2017. Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments. In *DIS 2017 - ACM Conference on Designing Interactive Systems (DIS '17 Proceedings of the 2017 Conference on Designing Interactive Systems)*. ACM, Edinburgh, United Kingdom, 1111–1122. https://doi.org/10.1145/3064663.3064695

[100] Erica Southgate, Shamus P. Smith, and Jill Scevak. 2017. Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth. In *2017 IEEE Virtual Reality (VR)*. 12–18. https://doi.org/10.1109/VR.2017.7892226

[101] Maximilian Speicher and Michael Nebeling. 2018. GestureWiz: A Human-Powered Gesture Design Environment for User Interface Prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173681

[102] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy. 2000. The gesture pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Digest of Papers. Fourth International Symposium on Wearable Computers*. 87–94. https://doi.org/10.1109/ISWC.2000.888469

[103] Stephen V. Stehman. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62, 1 (1997), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

[104] Kai Ming Ting. 2010. *Confusion Matrix*. Springer US, Boston, MA, 209–209. https://doi.org/10.1007/978-0-387-30164-8_157

[105] Laton Vermette, Shruti Dembla, April Y. Wang, Joanna McGrenere, and Parmit K. Chilana. 2017. Social CheatSheet: An Interactive Community-Curated Information Overlay for Web Applications. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 102 (dec 2017), 19 pages. https://doi.org/10.1145/3134737

[106] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the Bridge over Norman's Gulf of Execution: Revealing Feedforward's True Identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1931–1940. https://doi.org/10.1145/2470654.2466255

[107] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[108] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2018. Gesture Elicitation Studies for Mid-Air Interaction: A Review. *Multimodal Technologies and Interaction* 2, 4 (2018). https://doi.org/10.3390/mti2040065

[109] Gang Wang, Gang Ren, Xinye Hong, Xun Peng, Wenbin Li, and Eamonn O'Neill. 2022. Freehand Gestural Selection with Haptic Feedback in Wearable Optical See-Through Augmented Reality. *Information* 13, 12 (2022). https://doi.org/10.3390/info13120566

[110] Adam S. Williams, Jason Garcia, and Francisco Ortega. 2020. Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3479–3489. https://doi.org/10.1109/TVCG.2020.3023566

[111] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[112] Dennis Wolf, John J. Dudley, and Per Ola Kristensson. 2018. Performance Envelopes of in-Air Direct and Smartwatch Indirect Control for Head-Mounted Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 347–354. https://doi.org/10.1109/VR.2018.8448289

[113] Julia Woodward, Feben Alemu, Natalia E. López Adames, Lisa Anthony, Jason C. Yip, and Jaime Ruiz. 2022. "It Would Be Cool to Get Stampeded by Dinosaurs": Analyzing Children's Conceptual Model of AR Headsets Through Co-Design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 152, 13 pages. https://doi.org/10.1145/3491102.3501979

[114] Haijun Xia, Michael Glueck, Michelle Annett, Michael Wang, and Daniel Wigdor. 2022. Iteratively Designing Gesture Vocabularies: A Survey and Analysis of Best Practices in the HCI Literature. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 37 (may 2022), 54 pages. https://doi.org/10.1145/3503537

[115] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: Turning Your Cheek into an Input Surface by Embedded Optical Sensors on a Head-Mounted Display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 19, 8 pages. https://doi.org/10.1145/3139131.3139146

[116] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. 2008. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 193–202. https://doi.org/10.1109/ISMAR.2008.4637362

[117] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376233