Using Open-Ended Elicitation to Explore Adults' Virtual-Object Interaction Preferences in AR Headsets

Alejandro Delgado
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
alejandrodel@usf.edu

Md Mehedi Hasan Jibon
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
jibon@usf.edu

Hetvi Shah
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
hetvi@usf.edu

Fareeza Rahman
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
fareeza@usf.edu

Anzhelika Kurnikova
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
akurnikova@usf.edu

Julia Woodward
Bellini College of AI, Cybersecurity and Computing
University of South Florida
Tampa, Florida, USA
juliaevewoodward@usf.edu

Abstract

Current augmented reality (AR) headset applications often fail to consider users' natural interaction preferences, which leads to unintuitive interfaces and increased user frustration. We conducted an open-ended elicitation study with 30 adults to discover their natural interaction preferences in AR headsets. Participants proposed first-choice and second-choice interactions for 17 3D virtual cube actions (e.g., moving, rotating) using any interaction modality they preferred (e.g., gesture, speech, etc.). We broke down the participants' most common interactions for each cube action. We found that while hand gestures were the overwhelming first choice (82%), over half of all second choices (52%) involved a switch to a different modality (e.g., from gesture to speech), highlighting a critical need for interaction flexibility in AR headset applications. Additionally, our open-ended approach uncovered interactions that go beyond standard 3D commercial AR headset interactions, such as thinking, blowing air, and air-drawing. Based on our findings, we contribute an understanding of adults' natural AR headset expectations for 3D virtual object manipulation that can inform future applications.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; *Gestural input*.

Keywords

Augmented reality, gesture elicitation, open-ended elicitation, virtual object manipulation

ACM Reference Format:
Alejandro Delgado, Md Mehedi Hasan Jibon, Hetvi Shah, Fareeza Rahman, Anzhelika Kurnikova, and Julia Woodward. 2026. Using Open-Ended Elicitation to Explore Adults' Virtual-Object Interaction Preferences in AR Headsets. In *Proceedings of the 2026 International Conference on Advanced Visual Interfaces (AVI '26), June 08–12, 2026, Venice, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3811427.3811433

AVI '26, Venice, Italy

ACM ISBN 979-8-4007-2342-1/26/06
https://doi.org/10.1145/3811427.3811433

1 Introduction

Augmented reality (AR) is an immersive technology that displays virtual elements onto the real world [31]. AR headsets are becoming a popular choice to interact with AR environments as they provide increased user immersion and hands-free capabilities [21]. AR headsets are being used across multiple fields in which direct manipulation of 3D virtual objects is often an essential feature. For example, clinicians use AR headsets to manipulate 3D reconstructions of patient anatomy to rehearse procedures [7]. Similarly, in manufacturing, AR headsets enable technicians to use hand gestures to directly manipulate 3D holograms of mechanical workpieces to learn complex assembly sequences [1]. Using hand gestures is one of the most common ways to interact with AR headsets, as this approach intuitively mirrors how people engage with the real world [27]. Current AR headset applications often rely on an arbitrary set of gestures for interactions. However, due to the high variability of gesture preferences across users [17], arbitrary gesture sets frequently fail to match users' natural expectations, leading to unintuitive controls and frustration [14].

To combat these usability issues, researchers have started conducting gesture elicitation studies with AR headsets (e.g., [24, 25, 35]) to build more user-defined interaction sets. Although gesture elicitation has produced gesture sets that are more aligned with user expectations, prior AR headset elicitation studies constrained participants to a specific modality (e.g., gesture, speech) rather than letting participants choose a modality, limiting the analysis of participants' modality preferences and modality switching behavior. Additionally, the methodology utilized by previous elicitation studies is unable to capture participants' unique interaction proposals outside of standard modalities, preventing the study of participants' natural expectations of AR headset technology. A study by Delgado et al. [9] proposed an open-ended elicitation methodology for AR headsets, which aimed to address some of the aforementioned limitations. However, Delgado et al.'s methodology solely focused on analyzing the reasons behind user preferences (i.e., mental models), and did not quantify participants' proposed interactions. Therefore, there are still open research questions about users' modality preferences and natural expectations in the AR environment, which are critical for designing more intuitive AR headset applications. We expand on prior work by adapting an open-ended elicitation methodology to quantify participants' diverse interaction proposals and capture their natural modality preferences for foundational virtual-object manipulations in AR headsets.

We conducted an open-ended elicitation study with 30 adults in which they observed a 3D virtual cube perform 17 different actions (e.g., moving, rotating, getting bigger, creating, selecting, etc.) and proposed an interaction to replicate what they saw. We opted for a virtual cube as its simple shape allows for a clearer visual representation of actions such as rotation and scaling while avoiding the preconceived associations participants might have with more complex, recognizable objects. Additionally, prior AR headset elicitation studies have used a virtual cube [2, 9, 15, 35], making it an ideal choice to compare our results with prior work.

Our participants were free to use any interaction modality they preferred (e.g., gestures, body, speech). Similar to prior elicitation studies [9, 30], we collected a second interaction choice for each action, which allowed us to uncover a more holistic picture of participants' intuitive behaviors. After completing the study sessions, we quantified participants' interactions through qualitative coding, resulting in a codebook with 1,017 participant interactions. We classified participants' interactions into different categories of modalities (e.g., gesture, speech, body-based interaction), each encompassing multiple specific gestures or actions. We computed the Agreement Rates ($\mathcal{AR}$) and respective 95% bootstrapped confidence intervals for all referents, following prior work [32, 33, 37].

We analyzed our codebook to uncover participants' natural interaction preferences and overarching trends. We found that hand gestures were the overwhelmingly preferred first choice (82%), yet over half of all secondary proposals (52%) involved a switch to an entirely different modality (e.g., from gesture to speech) rather than a variation within the same modality. This pattern highlights both the diversity in participants' interactions and a strong preference for cross-modal flexibility. We also encountered many creative and unexpected proposals that emerged from our open-ended approach, such as using the mind to "think" the cube into existence or blowing air to move, rotate, or enlarge the cube. Our work expands current understanding of adults' natural AR headset expectations and provides open research directions and design recommendations for future AR headset elicitation studies and applications.

2 Related Work

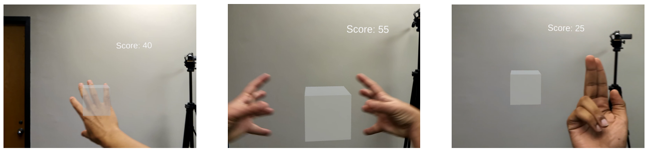
Gesture Elicitation is a study design in which participants are shown an *effect* (i.e., referent) and are asked to perform a gesture that would replicate the effect [36]. By collecting input directly from users, researchers can determine an appropriate set of gestures that will be easier to learn, recall, and perform for users [36, 37]. Gesture Elicitation has been used to improve various types of technologies such as whole-body interfaces [5, 20], mobile devices [11], tabletop computers [28, 37], as well as speech recognition [3]. Researchers have also started to use elicitation to study how users interact with AR headsets [15, 24, 25, 35, 38].

Piumsomboon et al. [25] conducted an AR headset elicitation study, in which participants observed a virtual task animation (e.g., shrinking virtual car) and designed gestures suited to perform the task. The authors reported that a majority of participants' gestures physically manipulated the virtual objects as if they were part of the real world. Pham et al. [24] conducted an AR headset elicitation study focused on analyzing users' interactions with three different-sized virtual objects: a miniature home model, a small city model, and a full-scale room model. The authors found that differences in the virtual object size and structure influenced participants' interaction behaviors, such as using fingertips to pick up small objects. Williams et al. [35] conducted an AR headset elicitation study, in which participants saw a virtual cube perform different actions (e.g., moving, expanding), and were prompted to propose one interaction per modality (i.e., gesture, speech, and gesture+speech), meaning each modality was elicited in isolation rather than chosen freely by participants. The authors found that the participants used gestures to interact with the cube as if it were a real object, and their proposed speech interactions were usually short commands. Li et al. [15] conducted a closed AR headset elicitation study in which participants were shown three pre-recorded videos, each demonstrating a different candidate gesture, and were asked to perform each demonstrated gesture. After replicating all three, participants selected the gesture they preferred most. The authors found a preference for simple one-handed gestures for manipulation tasks and one-handed pantomimic gestures for abstract tasks (e.g., create).

The aforementioned studies demonstrate the efficacy of elicitation methodology for discovering user expectations with different technologies. However, the methodology employed by these studies constrained participants to a single interaction modality at a time (e.g., only doing gestures or speech), which fails to capture whether users have specific modality preferences and whether those preferences remain consistent across referent groups, and restricts participants from proposing unique interactions that fall outside of standard modalities. In a recent elicitation study, Delgado et al. [9] worked to address the aforementioned limitations with the use of open-ended interaction prompts. However, Delgado et al. only analyzed participants' mental models for why they thought of certain interactions. We expand on prior work by utilizing an open-ended elicitation approach that can better capture unique interaction proposals, modality switching, and participants' overall interaction modality preferences across 17 virtual 3D cube actions.

3 Methodology

In our study, participants wore an AR headset and observed a 3D virtual cube perform 17 different actions (i.e., referents), such as moving upwards, and proposed an interaction to replicate what they saw (Figure 1). To determine our referents, we analyzed prior elicitation studies (e.g., [2, 9, 24, 35, 38]) and selected the 17 most prevalent referents that could be applied to a cube-shaped virtual object (Table 1). We chose a cube as its simple geometry communicates actions like rotation without predisposing participants to a specific type of interaction, and it allows us to better compare our results with prior AR headset elicitation studies [9, 35].



(A) (B) (C)

Figure 1: Participant's point-of-view from AR Headset in which they are: (a) grabbing and pushing the cube to *Move Right*; (b) expanding both hands to make the cube bigger (*Bigger*); (c) signaling the cube with two fingers to *Move Up.*

We classified our referents into three groups: (1) *Translation* for movements of the cube; (2) *Manipulation* for changes to the cube's properties (e.g., scale or rotation); and (3) *Abstract*, for higher-level or symbolic actions (Table 1). Our referent groupings were informed by prior work on interaction taxonomies with AR headsets [9, 35]. We counterbalanced the order of the three referent groups between participants to reduce ordering effects, while keeping the order of referents within each group fixed.

Table 1: List of the 17 referents used in our study.

Group	Referents
Translation	Move (Up/Down) Move (Left/Right) Move (Towards/Away)
Manipulation	Make (Bigger/Smaller) Rotate (X/Y/Z) axis Zoom (In/Out)
Abstract	Create Destroy Select (One/Multiple)

Participants first filled out a short demographic survey and completed a short practice with the AR headset to get familiar with the AR headset environment. During the practice, participants put on the AR headset, observed the static virtual cube, and were free to walk around while we verified equipment and headset fit. Our virtual cube spawned in front of the participant's eye line according to their height. To provide a standardized starting point for viewing each referent, we asked participants to stand on an X outlined on the floor while the referent was being shown. However, to encourage unconstrained movement and physical expression, we explicitly informed participants that after the referent was shown, they were free to move around to perform their proposed interaction. Each referent was played once initially; however, participants were allowed to ask the researcher to repeat the referent.

A goal of our methodology was to avoid constraining participants' proposals; therefore, we informed them at the beginning of the study that they could interact with the cube in any way they desired to replicate what they saw. For the main part of the study, participants would see the cube complete a referent, and we would ask them: *"How would you make the cube [current referent, e.g., Move Down] like you saw?"* for which they answered by performing an interaction (e.g., hand gesture, verbal command). We then asked: *"Why did you want to [Move Down] the cube like that?"* We utilized open-ended prompts to encourage participants to naturally explore their interaction and modality preferences.

After participants proposed their interaction and explained their reasoning, we asked them to rate their proposal from 1 to 5 based on *Goodness of Fit* ("The interaction I picked is a good match for its intended use") and *Ease of Use* ("The interaction I picked is easy to do"), which is consistent with prior work [9, 30, 37]. We then repeated the same process to record participants' second choice, a standard practice in gesture elicitation studies [9, 30]. Importantly, participants were free to propose a second interaction using the same modality as their first choice (e.g., two different hand gestures) or to switch to an entirely different modality (e.g., from gesture to speech). We did not prompt participants to change modalities, preserving the open-ended nature of the study. Each study lasted for around 60 minutes and was conducted in a spacious room that allowed for participants' unconstrained movement. Our study was also a within-subject study (i.e., participants saw all referents).

We recorded the participants' point-of-view from the headset, external audio, and external video. Our research protocol was approved by our Institutional Review Board, and each participant was compensated $40. We based our study design on the methodology used by Delgado et al. [9], as they focused on designing open-ended prompts to yield more natural responses from participants. However, while Delgado et al. focused on examining participants' mental models, our paper expands on their study design to allow a more in-depth analysis of participants' interaction proposals.

3.1 Participants

We recruited 30 adults between the ages of 18 and 54 (M=31.4, SD=12), 13 females and 17 males. A total of 27 participants self-reported they were right-handed, and 18 out of 30 had previous experience with VR headsets (e.g., Meta Quest, Oculus, Google Cardboard), typically from VR gaming (e.g., golf simulation, arcade games). We only had one person self-report AR headset experience from having volunteered for another AR research study in the past.

3.2 Equipment and Software

We used a *Magic Leap 2* [16] AR headset as it is lightweight and ideal for extended use. The headset has a FoV of 45H × 55V (70D) and a 1440 × 1760 resolution per eye. To display the cube, we created an application using *Unity* version *2022.3.22f1* and the Magic Leap 2 SDK [16]. We included a scoreboard in our application to increase engagement. The application spawned a virtual cube with 10-centimeter (cm) sides approximately 60 cm away from the front of the headset. All cube motions remained within the headset's FoV to keep the cube fully visible to the participant. Further details and referent animations are included in Supplementary Materials.

4 Data Analysis

4.1 Qualitative Coding

We analyzed participants' video-recorded interactions through qualitative coding (i.e., codebook model), a method for categorizing interaction properties based on a collaboratively developed set of descriptive codes [33]. Our codebook was informed by prior work

Table 2: Example codes used to describe each interaction modality. Code acronyms are represented in parenthesis, such as the code (PSH) representing the action Push. See the Supplementary Materials for full coding document.

Interaction Modality	Attribute	Example Codes
Gesture (G)	Number of Hands Arm Movement Hand Gestures Hand Actions	One Hand (1), Two Hands (2) Left (L), Right (R), Up (U), Down (D) One Finger (ONEF), Two Fingers (TWOF), etc. Push (PSH), Grab (GRB), Pinch (PIN), Craft (CRF), etc.
Speech (S)	Speech Commands	Single Word (SW), Multi Word (MW), etc.
Body (B)	Movement Direction	Approach (A), Retreat (RR), Left (L), Right (R)
Eye (E)	Movement Direction Facial Gestures	Eye Left (EL), Eye Right (ER), Eye Up (EU), Eye Down (ED) Eye Blink (EB), Eye Gaze (EG)
Head (H)	Movement Direction	Head Left (HL), Head Right (HR), Head Up (HU), etc.
Other (O)	Interaction Explanation	O:Thinking, O:Air Blow, etc.

on user-driven elicitation [26], multimodal interactions [35], and gesture elicitation techniques in augmented reality, virtual reality, and other technologies [10, 13, 24].

We coded participants' interactions based on six interaction modalities (i.e., Gesture, Speech, Body, Eye, Head, Other), with each modality having its own set of attributes (see Table 2 for more details): (1) *Gesture* interactions were defined as any movement of a bodily limb with the intention to complete the referent. We included codes for left and right arm movement, if they used one or two hands, as well as finger position (e.g., coding ONEF for a one-finger interaction) and intended action (e.g., coding "Push" (PSH) for a hand push) for each hand; (2) *Speech* interactions were defined as words/commands that come from the participant to complete the referent. We coded the verbatim utterance and the speech command type (single-word, multi-word, complete sentence); (3) *Body* interactions were defined as those in which the participant's body physically moved to another position to complete the referent; we also coded the direction of movement (e.g., coding "Approach" (A) if the participant approached the cube); (4) *Eye* interactions were coded when the participant specifically mentioned that they would use eye movement to complete the referent. We coded eye blink, gaze, and eye movement direction; (5) *Head* interactions were coded when the participant specifically mentioned that they would use their head to complete the referent; we also coded the head's movement direction; and (6) *Other* for when the interaction did not fit any aforementioned modality. Our coding approach allowed for capturing multimodal interactions (e.g., combining a *Gesture* with a *Speech* command; coded as G+S). For a full breakdown of all codes, see Supplementary Materials.

To establish a clear coding baseline, our research team first coded a participant (P1) together. Subsequently, the first five authors individually coded a non-overlapping subset of the remaining 29 participants (approximately 5–6 participants each). Throughout the coding process, we held weekly meetings to discuss edge cases, resolve disagreements through group consensus, and refine codebook definitions. When disagreements arose, we discussed the specific interaction until a unanimous interpretation was reached, and similar prior codes were revisited for consistency. Since our coding approach focused on group collaboration and discussion, we did not compute inter-rater reliability. Our iterative coding process allowed us to discuss and capture the semantic nuance of complex interactions that might be lost in a strictly independent coding protocol. Our initial codebook contained 1,020 potential entries (30 participants $\times$ 17 referents $\times$ 2 choices). We removed 3 entries due to missing data points, yielding 1,017 coded interactions. We then filtered out 10 interactions with a low average rating score (average $\leq$ 3) for the *Goodness of Fit* and *Ease of Use* statements (which all ended up being second-choice entries), as a low score indicated that the participant did not fully consider the interaction as a viable option. Our final dataset contained 1,007 interactions (510 first choices, 497 second choices).

4.2 Agreement Rates

After qualitatively coding the data, we used the Pandas Python library [22] to filter and process our data for quantitative analysis. We defined equivalence at the level of coded proposals: two proposals agreed if they matched on interaction modality (e.g., G, S, B, G+S) and action code for that modality (e.g., push, pull, grab, tap).

To quantify the degree of consensus among participants regarding their preferred interactions for each referent, we calculated the Agreement Rate ($\mathcal{AR}$) and its respective 95% bootstrapped confidence interval (see Table 3), as established by prior work [32, 33, 37]. Note that $\mathcal{AR}$ measures the consensus among participants' proposed interactions for a given referent, not the inter-coder reliability of the research team. We show the formal definition for a single referent r in Equation 1, where P is defined as the set of all interaction proposals for referent r, and P_i are the subsets of equivalent proposals from P. We also referenced the source code from the AGATe 2.0 tool (AGreement Analysis Toolkit) to compute $\mathcal{AR}$ [32]. Although the final set of codes observed in our data is finite, our codebook structure allowed for an infinite set of descriptive codes for each category (e.g., any novel interaction could emerge and be coded under "Other"). Following Vatavu and Wobbrock's [33] guidance, we therefore did not correct $\mathcal{AR}$ for chance agreement, as it assumes a finite, pre-defined set of response categories.

$$AR_r = \frac{\sum_{P_i \subseteq P} \frac{1}{2}|P_i|(|P_i| - 1)}{\frac{1}{2}|P|(|P| - 1)} \quad (1)$$

Table 3: Agreement Rate $\mathcal{AR}$ for first and second choices, with 95% bootstrapped confidence intervals (CIs). Cells are colored by agreement level (higher is better): red for low ($\leq$ .100), yellow for medium (.100−.300), green for high (.300−.500), and light-blue for very high (> .500) [32].

Referent	1st Choice $\mathcal{AR}$ (95% CI)	2nd Choice $\mathcal{AR}$ (95% CI)
Move Up	0.202 (0.124 – 0.379)	0.136 (0.101 – 0.264)
Move Down	0.269 (0.166 – 0.483)	0.110 (0.085 – 0.223)
Move Left	0.377 (0.239 – 0.595)	0.103 (0.080 – 0.223)
Move Right	0.292 (0.168 – 0.494)	0.113 (0.085 – 0.234)
Move Towards	0.315 (0.239 – 0.485)	0.189 (0.143 – 0.310)
Move Away	0.315 (0.207 – 0.517)	0.170 (0.115 – 0.326)
Destroy	0.115 (0.083 – 0.246)	0.115 (0.083 – 0.244)
Make Bigger	0.292 (0.221 – 0.437)	0.092 (0.076 – 0.191)
Make Smaller	0.280 (0.186 – 0.471)	0.140 (0.101 – 0.262)
Rotate (X)	0.260 (0.154 – 0.453)	0.163 (0.111 – 0.300)
Rotate (Y)	0.297 (0.182 – 0.501)	0.143 (0.097 – 0.280)
Rotate (Z)	0.398 (0.234 – 0.639)	0.239 (0.165 – 0.401)
Zoom In	0.166 (0.120 – 0.299)	0.117 (0.094 – 0.230)
Zoom Out	0.115 (0.090 – 0.228)	0.079 (0.071 – 0.170)
Create	0.189 (0.124 – 0.329)	0.186 (0.101 – 0.370)
Select One	0.294 (0.177 – 0.506)	0.080 (0.076 – 0.168)
Select Multiple	0.372 (0.230 – 0.591)	0.099 (0.078 – 0.205)
Overall Avg.	**0.267**	**0.134**

5 Results

Our analysis is based on 1,007 (510 first choices, 497 second choices) participant interactions. We first present the overall distribution of interaction modalities and then detail the specific interactions and agreement rates for each of the three referent groups (i.e., Translation, Manipulation, Abstract). See Table 4 for the modality distribution of participants' proposals across referent groups.

5.1 Overall Interactions

Gesture interactions were the most preferred modality, comprising the vast majority of responses (71.4%, $n = 719/1007$), followed by speech (20.6%, $n = 207/1007$). A majority of gestures (91.9%, $n = 661/719$) involved physically manipulating the cube. Participants had confidence in their interaction proposals, reflected in the high average ratings for *Goodness of Fit* ($M = 4.53/5$, $SD = 0.72$) and *Ease of Use* ($M = 4.61/5$, $SD = 0.65$). A Kruskal-Wallis test found no significant differences for *Goodness of Fit*, but revealed a significant divergence in *Ease of Use* for the Translation group ($H = 20.35, p = 0.005$). Descriptive statistics indicate this variance was driven by higher ratings for single-modality gestures ($M = 4.75, SD = 0.49$) compared to hybrid interactions such as Gesture+Eye or Gesture+Body ($M = 4.00, SD = 1.00$).

We observed a high degree of modality switching between first and second choices. Of the 497 instances in which participants provided two choices, 52.1% ($n = 259$) involved a switch to a different modality. The most dominant switching pattern was from gesture to speech, accounting for 50.6% ($n = 131/259$) of all switches, whereas the reverse switch occurred less frequently (19.7%, $n = 51/259$). Modality switching was consistent across referent groups, with Translation showing the highest switching rate at 53.9% ($n = 97/180$), followed by Manipulation (51.2%, $n = 103/201$) and Abstract (50.9%, $n = 59/116$).

Overall agreement for first choices was $\mathcal{AR}$=0.267, which is a medium level of agreement [32]; agreement decreased for second choices ($\mathcal{AR}$=0.134), as participants proposed more diverse interactions (e.g., walking, thinking, air blowing). Per-referent agreement rates and top two most common interactions are listed in Tables 3 and 5, respectively. In the following sections, we discuss referent-specific agreement rates and the interactions that drove them.

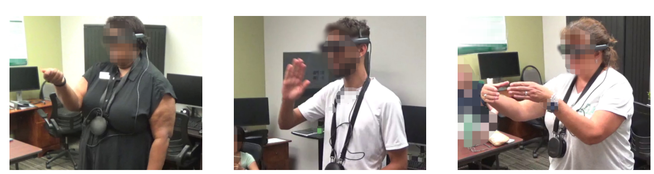
5.2 Translation Referent Group

For all translational referents, gestures were the dominant modality (70.6%, $n = 254/360$), followed by speech (18.1%, $n = 65/360$), most of which were second choices (72.3%, $n = 47/65$). We also saw an increased use of eye and head interactions for translation referents compared to the manipulation and abstract referents.

Both *Move Up* and *Move Down* had a medium level of agreement ($\mathcal{AR} = 0.202$ and $\mathcal{AR} = 0.269$). The most prevalent gesture for *Move Up* was an upwards pushing motion (50.0%, $n = 22/44$), followed by grabbing the cube and pushing it up (13.6%, $n = 6/44$). Three participants proposed symbolic gestures, such as drawing an upward arrow mid-air, pointing up to signal the cube (Figure 1), and fanning under the cube to generate an upward wind gust. For *Move Down*, the dominant gesture was a downwards pushing motion (50.0%, $n = 22/44$), followed by grabbing the cube and pushing it down (13.6%, $n = 6/44$). Two participants performed a mid-air phone-scrolling motion with their index finger to move the cube down. Speech followed an <Action> <Direction> <Distance> structure (e.g., "Move Up [X Distance]") and was predominantly a second choice for both referents.

Move Left showed the second-highest agreement of all referents ($\mathcal{AR}$=0.377), with a right-handed leftwards push as the dominant gesture (50.0%, $n = 21/42$), followed by grabbing and pushing left (21.4%, $n = 9/42$). *Move Right* also showed solid agreement ($\mathcal{AR}$=0.292), with a left-handed push in a rightwards direction being the dominant gesture (40.5%, $n = 17/42$). Speech for both referents followed the <Action> <Direction> <Distance> structure (e.g., "Move Left [X Distance]", "Move Right [X Distance]") and was predominantly a second choice.

Both *Move Towards* and *Move Away* elicited high agreement ($\mathcal{AR}$=0.315 each). For *Move Towards*, the most common gesture was grabbing the sides of the cube and pulling it towards the body (52.2%, $n = 24/46$), followed by pulling the cube from its back side (30.4%, $n = 14/46$). We also observed one participant using



(A) (B) (C)

Figure 2: Examples of participants' interactions, in which they are: (A) pulling an imaginary string to move the cube towards them (*Move Towards*); (B) simulating an erasing motion with their hand to *Destroy* the cube; (C) contouring the cube's shape (i.e., crafting) to *Create* the cube.

Table 4: Modality distribution of participants' proposals across referent groups. Combinations with only one proposal (E+H+B for *Rotate (X)*, G+H for *Move Right*, and S+H for *Select Multiple*) are omitted for brevity. Entries represent the total count of participant proposals (first choice, second choice) for each modality, with 60 entries per referent (30 participants × 2 choices). Referents with a total <60 are due to data filtering and/or missing data points.

Group	Referent	Total	Gesture (G)	Speech (S)	Body (B)	Eye (E)	Head (H)	G+B	E+H	Other (O)	G+E	G+S
Translation	Move Up	**60**	44 (26, 18)	9 (3, 6)	-	4 (1, 3)	1 (0, 1)	-	2 (0, 2)	-	-	-
	Move Down	**60**	44 (27, 17)	11 (2, 9)	-	1 (1, 0)	2 (0, 2)	-	2 (0, 2)	-	-	-
	Move Left	**60**	42 (27, 15)	10 (1, 9)	1 (0, 1)	2 (1, 1)	3 (0, 3)	-	1 (1, 0)	-	1 (0, 1)	-
	Move Right	**60**	40 (27, 13)	11 (1, 10)	1 (0, 1)	2 (1, 1)	4 (1, 3)	-	-	-	1 (0, 1)	-
	Move Towards	**60**	46 (24, 22)	11 (5, 6)	3 (1, 2)	-	-	-	-	-	-	-
	Move Away	**60**	38 (22, 16)	13 (6, 7)	2 (1, 1)	-	-	5 (1, 4)	-	1 (0, 1)	1 (0, 1)	-
Manipulation	Make Bigger	**60**	47 (27, 20)	12 (3, 9)	-	1 (0, 1)	-	-	-	-	-	-
	Make Smaller	**59**	47 (27, 20)	12 (3, 9)	-	-	-	-	-	-	-	-
	Rotate (X)	**58**	44 (26, 18)	11 (2, 9)	1 (1, 0)	-	-	1 (1, 0)	-	-	-	-
	Rotate (Y)	**60**	44 (29, 15)	10 (0, 10)	-	-	1 (1, 0)	2 (0, 2)	1 (0, 1)	2 (0, 2)	-	-
	Rotate (Z)	**57**	42 (27, 15)	14 (2, 12)	-	-	-	1 (1, 0)	-	-	-	-
	Zoom In	**59**	36 (20, 16)	16 (7, 9)	6 (3, 3)	-	-	-	-	-	1 (0, 1)	-
	Zoom Out	**58**	36 (18, 18)	13 (8, 5)	7 (3, 4)	-	-	1 (1, 0)	-	-	1 (0, 1)	-
Abstract	Create	**58**	38 (18, 20)	15 (11, 4)	-	1 (1, 0)	-	-	-	2 (0, 2)	-	2 (0, 2)
	Destroy	**60**	43 (24, 19)	16 (5, 11)	-	-	-	-	-	-	-	1 (1, 0)
	Select One	**60**	44 (26, 18)	11 (3, 8)	-	4 (1, 3)	-	-	-	-	-	1 (0, 1)
	Select Multiple	**58**	44 (25, 19)	12 (5, 7)	-	-	-	-	-	-	-	1 (0, 1)
Total		**1007**	**719** (420, 299)	**207** (67, 140)	**21** (9, 12)	**15** (6, 9)	**11** (2, 9)	**10** (4, 6)	**6** (1, 5)	**5** (0, 5)	**5** (0, 5)	**5** (1, 4)

a beckoning finger gesture and another miming the pulling of an imaginary string to move the cube towards them (Figure 2A). For *Move Away*, a pushing motion was the clear preference (71.1%, n = 27/38), followed by grabbing the cube with one hand before pushing it (13.2%, n = 5/38). Speech followed the <Action> <Direction> <Distance> structure (e.g., "Come Forward [X Distance]").

5.3 Manipulation Referent Group

Gestures were the preferred modality for all manipulation referents (72.0%, n = 296/411), followed by speech (21.4%, n = 88/411) and body interactions (3.4%, n = 14/411).

Agreement rates for making the object *Bigger* ($\mathcal{AR}$=0.292) and *Smaller* ($\mathcal{AR}$=0.280) were both high. The most common gesture for *Bigger* was a two-handed pulling-apart motion (34.0%, n = 16/47), typically using the index finger and thumb of each hand to pinch the corners of the cube and pull from opposite ends. Creative responses included one participant opening their eyes wide at the cube, another drawing a shaded cube outline with their finger to indicate a size increase, and one blowing air into the cube like blowing up a balloon. Speech (20.0%, n = 12/60) followed <Object> <Action> (e.g., "Cube Bigger") or <Action> <Measurement> (e.g., "Increase by 5%") structures. For *Smaller*, a pinching motion was most frequent (36.2%, n = 17/47), followed by pushing the sides of the cube together with both hands (23.4%, n = 11/47). Interestingly, one response involved using the right hand to mimic erasing part of the cube to make it appear smaller (Figure 2B). Speech was predominantly a second choice for both referents.

The rotational referents elicited the highest agreement rates in the study, likely due to participants drawing on gestures for rotating real-world objects (Figure 3). *Rotate Z* achieved the highest agreement of all referents ($\mathcal{AR}$=0.398), followed by *Rotate Y* ($\mathcal{AR}$=0.297) and *Rotate X* ($\mathcal{AR}$=0.260). For *Rotate X*, 52.3% (n = 23/44) of gestures involved grabbing the bottom of the cube and performing a clockwise turn. We saw a unique body positioning strategy for *Rotate X*, in which one participant walked in a semi-circle around the cube to replicate the cube's horizontal rotation. For *Rotate Y*, the most common gesture was grabbing the side of the cube with one hand and rotating it vertically away from the body (45.7%, n = 21/44). We observed two instances in which participants blew air towards the cube to tilt the cube backwards. For *Rotate Z*, the preferred gesture was a grab and counter-clockwise turn gesture (61.9%, n = 26/42), similar to loosening a valve. Speech followed the <Action> <Direction> structure (e.g., "Rotate Clockwise") and was largely a second choice for all rotation referents (see Table 4).

In contrast to the rotation tasks, zoom referents showed a lower consensus, with $\mathcal{AR}$=0.166 for *Zoom In* (i.e., cube gets bigger and moves closer) and $\mathcal{AR}$=0.115 for *Zoom Out*. The dominant gestures generally used two or more fingers, with a pinch-out motion for *Zoom In* (41.7%, n = 15/36) and a pinch-in motion for *Zoom Out* (33.3%, n = 12/36). We also observed increased body modality use, with participants walking closer to the cube to *Zoom In* (n = 6) and away to *Zoom Out* (n = 7). Speech followed an <Action> <Direction> format (e.g., "Zoom In", "Zoom Out").

5.4 Abstract Referent Group

While gestures remained the primary choice for abstract referents (71.6%, n = 169/236), speech saw its highest use out of all referent groups (22.9%, n = 54/236).

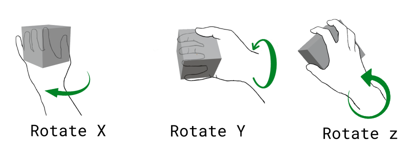


Figure 3: Figure illustrations for the most common gesture used for the *Rotate X/Y/Z* referents.

Table 5: Most common first and second proposed interactions for each referent. Multiple interaction proposals with equal occurrence are separated by a vertical bar symbol (|). MW denotes multi-word speech commands.

Referent	First Most Common Interaction	Second Most Common Interaction
Move Up	Upward push	Speech MW (e.g., "Move Up [X Distance]")
Move Down	Downward push	Speech MW (e.g., "Cube Move Down")
Move Left	Leftward push	Grab/leftward push | Speech MW (e.g., "Move Left [X Distance]")
Move Right	Rightward push	Speech MW (e.g., "Move Right [X Distance]")
Move Towards	Grab and pull	Pull | Speech MW (e.g., "Come Forward [X Distance]")
Move Away	Forward push	Speech MW (e.g., "Move Away [X Distance]")
Make Bigger	Pull apart with two hands	Pinch-out
Make Smaller	Pinch-in	Pushing opposite sides together (i.e., squishing)
Rotate (X)	Grab and rotate clockwise	Speech MW (e.g., "Rotate Left [X Degrees]")
Rotate (Y)	Grab and rotate away from the body	Speech MW (e.g., "Rotate Up")
Rotate (Z)	Grab and rotate counter-clockwise	Speech MW (e.g., "Rotate Counter Clockwise")
Zoom In	Pinch-out	Speech MW (e.g., "Zoom In")
Zoom Out	Pinch-in	Speech MW (e.g., "Zoom Out")
Create	Craft (i.e., contouring cube's outline)	Speech MW (e.g., "Create a Cube")
Destroy	Forceful push (i.e., punching)	Speech MW (e.g., "Remove Cube")
Select One	Tap	Speech MW (e.g., "Select Cube")
Select Multiple	Tap	Speech MW (e.g., "Select Cube 1, Select Cube 3")

Create elicited a diverse set of interactions, resulting in a relatively low agreement rate ($\mathcal{AR}$=0.189). The most prevalent gesture involved using two hands to "craft" or outline the shape of the 3D cube in mid-air (55.3%, $n = 21/38$) (Figure 2C), followed by finger snapping (5.3%, $n = 2/38$) and two instances of drawing a 2D square in mid-air using one finger. Interestingly, we observed two instances of "thinking" the cube into existence. Speech (25.9%, $n = 15/58$) followed an <Action> <Object> structure (e.g., "Create a Cube"). For *Create*, $n = 11/15$ of the speech interactions (73.33%) were performed during the participants' first-choice proposal.

Destroy had one of the lowest agreement rates ($\mathcal{AR}$=0.115), with a preference for a forceful pushing/punching motion (41.9%, $n = 18/43$), followed by finger snapping (7.0%, $n = 3/43$). Creative gestures included erasing the cube similar to erasing a whiteboard ($n = 2$), drawing an X in mid-air ($n = 2$), forming an X with their arms ($n = 1$), flicking the cube away ($n = 2$), and swiping the cube away ($n = 2$). *Destroy* saw the highest speech use of all referents (26.7%, $n = 16/60$) following the <Action> <Object> structure (e.g., "Delete the Cube") and being predominantly a second choice.

In contrast, selecting objects showed strong agreement. *Select One* ($\mathcal{AR}$=0.294) and *Select Multiple* ($\mathcal{AR}$=0.372) both demonstrated high consensus, with index finger tapping as the preferred gesture (45.5%, $n = 20/44$ and 50.0%, $n = 22/44$, respectively). One participant turned their head toward each cube to select them. Speech (18.3% for *Select One*; 20.7% for *Select Multiple*) followed an <Action> <Object> syntax (e.g., "Select Cube 1, Cube 3").

6 Discussion

Consistent with prior AR headset elicitation studies with both adults and children [8, 9, 24, 25], the majority of our participants preferred gestures that physically manipulated the virtual cube as if it were a real object (i.e., direct manipulation). This finding can be explained through Hutchins et al.'s [12] direct manipulation framework. Hutchins et al. studied two underlying phenomena that elicit a feeling of directness in users: (1) minimizing the cognitive distance between the user's intention and the action required by the system, and (2) providing object representations that behave as if they are the objects themselves. Both principles were reflected in Delgado et al.'s [9] mental model analysis of AR headset interactions, in which participants chose interactions they believed required less effort and expected the virtual cube to respond to physical forces like a real-world object. We acknowledge that the spatial nature of our referents (e.g., translation, rotation, scaling) likely contributed to gesture dominance as a first-choice modality, as spatial referents naturally exhibit low cognitive distance between intention and action when performed through gesture [29]. However, this preference was not absolute, as our results revealed that even for inherently spatial referents, a meaningful subset of participants (13.1%) preferred speech.

While gesture dominance was broadly consistent with prior work, we found significant differences in the form of proposed interactions for abstract referents. For the *Create* referent, Williams et al. [35] and Zhou et al. [38] both reported the most elicited interaction to be a "bloom" gesture, which is also the standard gesture for opening menus in the Microsoft HoloLens 1 [18]. We found that none of our participants elicited the bloom gesture for *Create*, instead opting to craft the cube (i.e., contouring cube's outline). This discrepancy is likely due to the fact that only 1 of our participants (out of 30) had used an AR headset before, compared to 45.8% (n=11/24) and 12.5% (n=3/24) of the participants in Williams et al.'s and Zhou et al.'s studies. For *Destroy*, our participants preferred a forceful pushing/punching motion, which contrasts with Williams et al.'s reported swiping gesture and Zhou et al.'s pointing gesture.

We found that proposed speech commands followed a simple syntactic structure (e.g., <Action> <Direction> <Distance> for *Move Up*), similar to prior work [35, 38]. However, in our study, speech commands for abstract referents had an increased mention of the <Object> parameter, which differs from Williams et al. [35], who found that the <Object> parameter in speech commands was a rare occurrence. The explicit reference of the object could be beneficial for abstract tasks that involve the creation or selection of an object.

Prior AR headset elicitation studies [35, 38] explicitly prompted participants to propose Gesture+Speech interactions. In contrast, Gesture+Speech interactions were largely absent in our open-ended study (0.5%, $n = 5/1007$), appearing only in the abstract referent group. This suggests that while users can perform multimodal interactions when prompted, they do not intuitively default to them when given freedom. This behavior aligns with Khurana et al. [14], who found that novice AR headset users often lack robust mental models for AR interactions, relying instead on legacy metaphors (e.g., 2D touchscreens) that typically favor unimodal input. Consequently, the absence of Gesture+Speech in our study likely reflects a discoverability gap rather than a lack of utility. Future work should **research design strategies that can connect users' legacy mental models with AR's multimodal capabilities.**

Our open-ended elicitation approach allowed for participants to propose a range of interactions (e.g., body, head, eyes) that previous AR headset elicitation studies were unable to capture due to modality and interaction constraints in their methodology. This interaction and modality freedom led to participants proposing a variety of unique and interesting interactions, such as using their hand to "fan under the cube" to generate an upwards wind gust that would move the cube up, blowing air on the cube to make it bigger, using one finger to draw a 2D square in mid-air to create the cube, or using their minds to "think" the cube into existence. Our findings highlight that participants often think beyond the standard interactions offered by commercial AR headsets (e.g., grab virtual objects using air tap for the Microsoft HoloLens 2 [19]) and expect different interactions with the technology. **We recommend the adoption of an open-ended methodology for future elicitation studies, as it better captures participants' natural expectations.**

We found that participants consistently used a variety of modalities (gestures, speech, eye, head, body, etc.) to interact with the virtual cube across all referent groups. Furthermore, over half of all second-choice proposals (52%) involved a switch to a different modality, indicating that, as a fallback mechanism, participants often preferred switching modalities rather than proposing an alternative interaction within the same modality. Therefore, **future AR headset applications should accommodate interaction flexibility to better align with users' expectations.**

An important consideration when asking for a second-choice interaction for each referent is whether it may have biased participants toward proposing a deliberately different modality. Our quantitative results showed that 47.9% of second-choice proposals remained within the same modality as the first choice, indicating that participants did not uniformly default to switching modalities. We believe that the observed switching rate more likely reflects a cross-modal fallback preference than a methodological artifact.

In further analysis of modality switching behavior, we found that speech was proposed more frequently as a primary modality for abstract referents. This increased use of speech for abstract referents is consistent with the direct manipulation framework [12], as abstract tasks such as *Create* and *Destroy* do not map naturally onto physical actions the way spatial tasks do. More broadly, Cha et al. [4] investigated how usage context influences users' modality selection (i.e., gesture and speech) in three different contexts (reading, watching, driving). The authors identified the existence of a threshold of physical and mental effort in which users start to prefer one modality over the other, with the threshold amount varying according to context. Our participants' preference for modality switching and increased use of speech for abstract referents in AR headsets outlines **future research directions to determine what contexts and factors trigger modality switching.**

7 Limitations & Future Work

There are limitations to the scope of our work. Although we had a good balance between the number of female ($n = 13$) and male ($n = 17$) participants, we did not analyze whether there were any interaction differences based on gender or other demographic features. Future work should investigate whether variables such as gender, age, or cultural background influence modality preferences. While 30 participants may seem small, it is consistent with prior work [6, 9, 20, 23, 34, 35]. Participants viewed a 3D virtual cube performing 17 simple referents in a quiet lab setting; real-world AR introduces environmental factors (e.g., ambient noise, social context, crowd density) that may affect perceived modality suitability. Future work should examine more complex referents and objects for different interaction contexts and environmental factors. Furthermore, we did not examine how AR-specific perceptual factors (e.g., FoV constraints, occlusion, depth perception) could influence users' expectations. Future work could analyze if these factors significantly influence users' expectations in AR headsets. In addition, future work can evaluate the participants' proposed interactions in terms of usability, effectiveness, and cognitive workload.

8 Conclusion

Although AR headsets are increasingly being used in a wide range of contexts (e.g., education, medicine), AR headset applications often fail to consider users' natural interaction preferences, which can increase user frustration. We conducted an open-ended AR headset elicitation study with 30 adults in which they proposed interactions for 17 virtual cube actions (e.g., moving, rotating) using any interaction modality they preferred (hands, body, voice, etc.). Compared to prior work, we observed more variation in participants' proposed interactions (e.g., blowing on the cube, eye/head/body movement). While hand gestures were the preferred first choice (82%), over half of all second choices (52%) involved a switch to a different modality, underscoring the importance of designing AR interfaces that support seamless modality switching to accommodate user preferences, accessibility needs, and situational constraints. We also found that speech use increased for abstract referents in which gesture loses its articulatory advantage, suggesting that modality preferences could be task-dependent rather than fixed. We contribute an empirical foundation of adults' natural AR interaction preferences and a set of design recommendations grounded in these preferences.

References

[1] George Apostolopoulos, Dionisis Andronas, Nikos Fourtakas, and Sotiris Makris. 2022. Operator Training Framework for Hybrid Environments: An Augmented Reality Module Using Machine Learning Object Recognition. *Procedia CIRP* 106 (Jan. 2022), 102–107. doi:10.1016/j.procir.2022.02.162

[2] 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 (May 2020), 214–228. doi:10.1080/15230406.2019.1696232

[3] Nabil Al Nahin Ch, Diana Tosca, Tyanna Crump, Alberta Ansah, Andrew Kun, and Orit Shaer. 2022. Gesture and Voice Commands to Interact With AR Windshield Display in Automated Vehicle: A Remote Elicitation Study. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 171–182. doi:10.1145/3543174.3545257

[4] Min Chul Cha and Yong Gu Ji. 2024. Context Matters: Understanding the Effect of Usage Contexts on Users' Modality Selection in Multimodal Systems. *International Journal of Human–Computer Interaction* 40, 20 (Oct. 2024), 6287–6302. doi:10.1080/10447318.2023.2250606

[5] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz Elicitation Study Examining Child-Defined Gestures with a Whole-Body Interface. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*. Association for Computing Machinery, New York, NY, USA, 277–280. doi:10.1145/2485760.2485823

[6] Andreea Danielescu and David Piorkowski. 2022. Iterative Design of Gestures During Elicitation: Understanding the Role of Increased Production. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3491102.3501962

[7] Lucio Tommaso De Paolis, Sergio Teodoro Vite, Miguel Ángel Padilla Castañeda, César Fabian Domínguez Velasco, Salvatore Muscatello, and Aldo Francisco Hernández Valencia. 2022. An Augmented Reality Platform with Hand Gestures-Based Navigation for Applications in Image-Guided Surgery: Prospective Concept Evaluation by Surgeons. *International Journal of Human–Computer Interaction* 38, 2 (Jan. 2022), 131–143. doi:10.1080/10447318.2021.1926116

[8] Alejandro Delgado, Md Mehedi Hasan Jibon, Hetvi Shah, Fareeza Rahman, Anzhelika Kurnikova, and Julia Woodward. 2026. "If you are a Star Wars fan, use the force": Exploring Children's Virtual-Object Interaction Preferences in AR Headsets. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, Barcelona, Spain, 1–18. doi:10.1145/3772318.3790552

[9] Alejandro Delgado, Hetvi Shah, Md Mehedi Hasan Jibon, Anzhelika Kurnikova, Fareeza Rahman, and Julia Woodward. 2025. "Because We Are Human, We Use Hands": Understanding People's Expectations and Mental Models of Virtual Object Interaction in AR Headsets. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3706599.3719691

[10] Haiwei Dong, Ali Danesh, Nadia Figueroa, and Abdulmotaleb El Saddik. 2015. An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions. *IEEE Access* 3 (2015), 543–555. doi:10.1109/ACCESS.2015.2432679

[11] Shariff A. M. Faleel, Michael Gammon, Yumiko Sakamoto, Carlo Menon, and Pourang Irani. 2020. User Gesture Elicitation of Common Smartphone Tasks for Hand Proximate User Interfaces. In *Proceedings of the 11th Augmented Human International Conference (AH '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3396339.3396363

[12] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct Manipulation Interfaces. *Hum.-Comput. Interact.* 1, 4 (Dec. 1985), 311–338. doi:10.1207/s15327051hci0104_2

[13] Sumbul Khan and Bige Tunçer. 2019. Gesture and Speech Elicitation for 3D CAD Modeling in Conceptual Design. *Automation in Construction* 106 (Oct. 2019), 102847. doi:10.1016/j.autcon.2019.102847

[14] Anjali Khurana, Michael Glueck, and Parmit K. Chilana. 2024. Do I Just Tap My Headset? How Novice Users Discover Gestural Interactions with Consumer Augmented Reality Applications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4 (Jan. 2024), 165:1–165:28. doi:10.1145/3631451

[15] Jieqiong Li, Adam S. Coler, Zahra Borhani, and Francisco R. Ortega. 2024. Collecting and Analyzing the Mid-Air Gestures Data in Augmented Reality and User Preferences in Closed Elicitation Study. In *Virtual, Augmented and Mixed Reality*, Jessie Y. C. Chen and Gino Fragomeni (Eds.). Springer Nature Switzerland, Cham, 201–215. doi:10.1007/978-3-031-61044-8_15

[16] leap magic. [n. d.]. Magic Leap 2: Next-Gen Augmented Reality Headset for Professionals. https://www.magicleap.com/magic-leap-2.

[17] David McNeill. 2005. *Gesture and Thought*. University of Chicago Press, Chicago, IL, US. xii, 318 pages. doi:10.7208/chicago/9780226514642.001.0001

[18] Microsoft. [n. d.]. HoloLens (1st Gen) Hardware. https://learn.microsoft.com/en-us/previous-versions/mixed-reality/hololens-1/hololens1-hardware.

[19] Microsoft. 2025. Getting around HoloLens 2. https://learn.microsoft.com/en-us/hololens/hololens2-basic-usage.

[20] Francisco R. Ortega, Katherine Tarre, Mathew Kress, Adam S. Williams, Armando B. Barreto, and Naphtali D. Rishe. 2019. Selection and Manipulation Whole-Body Gesture Elicitation Study In Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1723–1728. doi:10.1109/VR.2019.8798105

[21] Ahmed Oun, Nathan Hagerdorn, Caleb Scheideger, and Xiangyi Cheng. 2024. Mobile Devices or Head-Mounted Displays: A Comparative Review and Analysis of Augmented Reality in Healthcare. *IEEE Access* 12 (2024), 21825–21839. doi:10.1109/ACCESS.2024.3361833

[22] The pandas development team. 2020. *pandas-dev/pandas: Pandas*. doi:10.5281/zenodo.3509134

[23] Jorge-Luis Pérez-Medina, Santiago Villarreal, and Jean Vanderdonckt. 2020. A Gesture Elicitation Study of Nose-Based Gestures. *Sensors* 20, 24 (Jan. 2020), 7118. doi:10.3390/s20247118

[24] 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 (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 227–240. doi:10.1145/3196709.3196719

[25] 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, 282–299. doi:10.1007/978-3-642-40480-1_18

[26] Shwetha Rajaram, Chen Chen, Franziska Roesner, and Michael Nebeling. 2023. Eliciting Security & Privacy-Informed Sharing Techniques for Multi-User Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3581089

[27] Kasper Rise and Ole Andreas Alsos. 2020. The Potential of Gesture-Based Interaction. In *Human-Computer Interaction. Multimodal and Natural Interaction*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 125–136. doi:10.1007/978-3-030-49062-1_8

[28] Karen Rust, Meethu Malu, Lisa Anthony, and Leah Findlater. 2014. Understanding Childdefined Gestures and Children's Mental Models for Touchscreen Tabletop Interaction. In *Proceedings of the 2014 Conference on Interaction Design and Children (IDC '14)*. Association for Computing Machinery, New York, NY, USA, 201–204. doi:10.1145/2593968.2610452

[29] Ben Shneiderman. 1981. Direct Manipulation: A Step beyond Programming Languages (Abstract Only). In *Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems. (Part - II): Human Interface and the User Interface - Volume 1981 (CHI '81)*. Association for Computing Machinery, New York, NY, USA, 143. doi:10.1145/800276.810991

[30] Nikita Soni, Schuyler Gleaves, Hannah Neff, Sarah Morrison-Smith, Shaghayegh Esmaeili, Ian Mayne, Sayli Bapat, Carrie Schuman, Kathryn A. Stofer, and Lisa Anthony. 2019. Do User-Defined Gestures for Flatscreens Generalize to Interactive Spherical Displays for Adults and Children?. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis '19)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3321335.3324941

[31] D.W.F. Van Krevelen and R. Poelman. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality* 9, 2 (Jan. 2010), 1–20. doi:10.20870/IJVR.2010.9.2.2767

[32] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1325–1334. doi:10.1145/2702123.2702223

[33] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2022. Clarifying Agreement Calculations and Analysis for End-User Elicitation Studies. *ACM Trans. Comput.-Hum. Interact.* 29, 1 (Jan. 2022), 5:1–5:70. doi:10.1145/3476101

[34] 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 (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. doi:10.1145/3357236.3395511

[35] 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 (Dec. 2020), 3479–3489. doi:10.1109/TVCG.2020.3023566

[36] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1869–1872. doi:10.1145/1056808.1057043

[37] 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 (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1083–1092. doi:10.1145/1518701.1518866

[38] Xiaoyan Zhou, Adam Sinclair Williams, and Francisco Raul Ortega. 2022. Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3562939.3565637