User-Defined Gestures with Physical Props in Virtual Reality

MARCO MORAN-LEDESMA, University of Waterloo, Canada
OLIVER SCHNEIDER, University of Waterloo, Canada
MARK HANCOCK, University of Waterloo, Canada

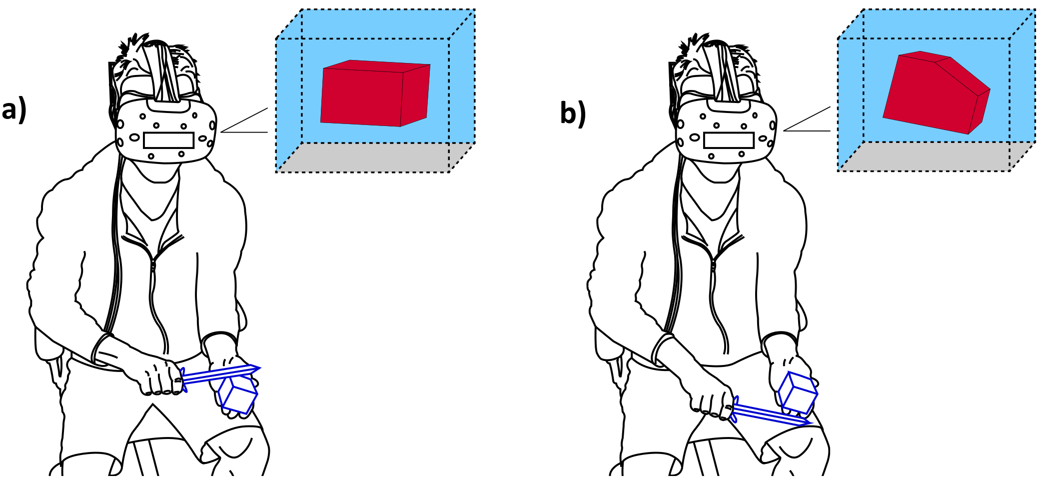


Fig. 1. We elicited manipulative gestures from 21 participants to explore how they might use physical props to control abstract actions in VR, such as beveling an object.

When interacting with virtual reality (VR) applications like CAD and open-world games, people may want to use gestures as a means of leveraging their knowledge from the physical world. However, people may prefer physical props over handheld controllers to input gestures in VR. We present an elicitation study where 21 participants chose from 95 props to perform manipulative gestures for 20 CAD-like and open-world game-like referents. When analyzing this data, we found existing methods for elicitation studies were insufficient to describe gestures with props, or to measure agreement with prop selection (i.e., agreement between sets of items). We proceeded by describing gestures as context-free grammars, capturing how different props were used in similar roles in a given gesture. We present gesture and prop agreement scores using a generalized agreement score that we developed to compare multiple selections rather than a single selection. We found that props were selected based on their resemblance to virtual objects and the actions they afforded; that gesture and prop agreement depended on the referent, with some referents leading to similar gesture choices, while others led to similar prop choices; and that a small set of carefully chosen props can support multiple gestures.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: Virtual Reality; 3D physical props; Gestural input; Elicitation technique; Agreement score; Similarity measures, Immersive interaction.

Authors' addresses: Marco Moran-Ledesma, Systems Design Engineering, University of Waterloo, Waterloo, Canada, mamoranl@uwaterloo.ca; Oliver Schneider, Management Sciences, University of Waterloo, Waterloo, Canada, oliver.schneider@uwaterloo.ca; Mark Hancock, Management Sciences, University of Waterloo, Waterloo, Canada, mark.hancock@uwaterloo.ca.

ACM Reference Format:
Marco Moran-Ledesma, Oliver Schneider, and Mark Hancock. 2021. User-Defined Gestures with Physical Props in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 5, ISS, Article 488 (November 2021), 23 pages. https://doi.org/10.1145/3486954

1 INTRODUCTION

In virtual reality (VR) applications, like computer-aided design (CAD) or games, people may want to use gestures, which provide an interaction input that is often described as "natural" and "intuitive" [31], as well as allowing for easy spatial manipulation [2]. However, rather than invoking gestures for CAD and games with VR controllers, people may prefer using physical objects (props), which are effective at improving immersion [6] and enhancing realism [26, 56]. We propose to expand the input vocabulary of gestures by combining them with physical props.

To design gestures, researchers have widely adopted the elicitation approach [42, 71], where people who might use the system provide their input. These studies have shown promising outcomes for gesture design in various domains, for example, tabletops [71], public displays [35], mobile platforms [23, 37, 54], keyboards [7], tangible systems [9, 66], smartwatches [3], virtual reality [45], and augmented reality [50]. Elicited gestures have been shown to be easier to remember and preferred by those without technical expertise [44].

In our research, we conduct an elicitation study for manipulative gestures [32, 51] with physical props in VR. We chose referents from CAD modelling software and open world games to ground our work in application areas that offer a rich vocabulary of actions and commands. We followed Wobbrock et al.'s approach [71] by first showing the effect of an action (called a *referent*) in VR and then asking participants for their preferred gesture, but also asking them to choose their preferred prop(s) to perform the gesture with. Our results include a set of twenty user-defined gestures with props, one per referent.

However, when analyzing our data, we found existing methods were insufficient. Specifically, we had no effective way to describe gestures with props, because people often chose different props for what appeared to be the same gesture (e.g., both did a "cutting" action, one with a knife, another with a sword). We also had no way to calculate agreement between people for prop selection, because current methods did not consider partial similarity when participants chose multiple simultaneous props (e.g., two participants choosing one prop in common and another that differed).

To analyze our data, we developed two representations to handle both gestures and props. 1) We found that context-free grammars were a useful method to capture the various props used in a gesture, their role, and how people used them; they are presented alongside illustrations of our elicited gestures. This language succinctly communicates gestures, and could be directly implemented into systems using props for gestures, and can support future, more-involved analysis. 2) We calculated agreement scores for gestures using methods from previous elicitation studies [68, 71], but these scores were inadequate for the multiple props that participants selected. We thus introduce a new agreement score based on set similarity metrics, and use it to analyze agreement between both gestures and props. This score is identical to previous agreement scores [68] when used for a single selection, but also accommodates multiple selections.

Our contributions are: (1) a set of elicited gesture-prop combinations for 20 CAD-like and open–world-like referents; (2) a language for articulating gestures with props, based on context-free grammars; (3) a generalization of the agreement metric to account for multiple selections in elicitation studies; and (4) design recommendations for gestures involving props in VR environments.

We do not implement the proposed gesture set in a working VR system in this paper, but our work is a necessary step toward this implementation, and provides a deeper understanding of how people interact with props to complete actions in VR (Figure 2). We explore what props are chosen and what gestures are depicted with them, and discuss how these gestures could be realized and what interaction designers can learn from our findings.

2 RELATED WORK

Relevant work includes an overview of previous work around gesture elicitation studies in various domains, gesture-based user interfaces for CAD modelling, adoption of VR for CAD 3D modelling, and haptic technology in VR.

2.1 Gesture Elicitation Studies in HCI

Gesture elicitation is a widely used technique in Human-Computer Interaction (HCI) for identifying gesture vocabularies that are self-discoverable [65]. Wobbrock et al. [71] developed a user-defined set of gestures based on the degree of consensus (agreement score [72]) among participants to complete 27 referents. They also classified the elicited vocabulary of gestures in a taxonomy for tabletop systems design, which aims to capture the gesture design space in a tabletop environment.

Since that work, elicitation studies have become a common practice for determining suitable gestures, but so far such studies have been limited to actions using hands and fingers with specific technology (e.g., public displays [35], mobile platforms [23, 37, 54], televisions [67], keyboards [7], tangible interfaces [9, 66], smartwatches [3], augmented and virtual reality [45, 50]). In our work, we adopt the elicitation study methodology to determine a gesture set for using physical props to control virtual reality, and by necessity build on this previous work to be able to incorporate not only the choice of gesture, but also the choice of physical artifact used with the gesture.

2.2 Gesture-based User Interfaces for CAD 3D Modelling

In our work, we explore the domain of 3D modelling in computer-automated design (CAD), and some work has already explored the use of gestures to perform 3D modelling tasks. Khan and Tunçer [34] presented a compilation of a set of gestures and speech commands for 3D CAD modelling for conceptual design that were elicited from participants and evaluated by experts individually. In this study, the authors included modelling tasks like “rotate”, “scale”, and “zoom in”, which are some of the referents in this paper. Huang and Rai [29] also presented a system that recognized hand gestures along with hand position information and converted them into commands for rotating, translating and scaling 3D models. We expand the vocabulary of commands (referents) by incorporating tasks such as “changing colour”, “bending”, “perforating”, “beveling”. More recently, De Araújo et al. [18] implemented *Mockup Builder*, a 3D modelling system combining gestures on a multi-touch surface and gestures in 3D space to do CAD modelling tasks. Mockup builder is a stereoscopic display where users can directly interact with 3D models or edit them using gestures on or above the surface. We extracted referents like “extruding a model” and “reshaping a prism” from this work.

2.3 Virtual Reality for CAD 3D Modelling

There has been an extensive amount research that employs VR and/or AR to do CAD modelling. We focus our discussion on VR, rather than AR systems like *DesignAR* [53].

Both research and commercial systems have proposed systems whose main input device is one or more handheld controllers. Feeman et al. [20] developed a platform for testing CAD in VR through interaction with a game engine (Autodesk’s Stingray), in which people modelled chairs, trucks, and mazes with a Vive controller. McGraw et al. [40] proposed an interactive technique that enables artists and designers to create sculptural forms and spatial surfaces using two Vive controllers. Currently, Autodesk’s Create VR [5] design tool lets artists and designers immersively explore 3D models with either Vive or Oculus controllers.

Other researchers have customized their own handheld controllers. Mine et al. [41] built a controller that combined a smartphone and a casing with physical buttons to interact with an adapted VR version of the SketchUp modelling software. Butterworth et al. [10] developed *3dm*, a three-dimensional surface modelling interface where users can model primitive shapes and perform CAD-like actions using a 6D 2-button mouse. Jackson and Keefe [30] presented *Lift-Off*, an immersive 3D interface for creating complex models with stylus pens, which feels natural to artists. Keefe et al. [30] introduced *Drawing on Air*, an input technique to draw 3D lines and curves using a 6 DOF tracker and the Phantom haptic device. This system is thus haptic-aided.

In our work, rather than using customized or existing handheld controllers for interaction, we propose using passive objects and manipulative gestures. CAD modelling (mainly) and open-world games serve as our test beds for this form of interaction.

2.4 Haptics Technology in Virtual Reality

There has been an impressive amount of research in the area of incorporating physical objects or devices and haptic feedback in virtual reality experiences, in the form of active haptics, passive haptics, and dynamic passive haptics.

2.4.1 Active Haptics in Virtual Reality. Active haptics provide touch feedback using a variety of actuation methods. Researchers in this area have spent a significant amount of time and effort customizing controllers for input in VR, such as grounded/ungrounded shape-changing surfaces [1, 61], electro-mechanical actuators [16, 60], pneumatically-actuated interfaces [19, 64], and devices rendering touch and texture [8, 70]. These approaches offer interaction techniques that realistically simulate how people interact with physical objects in the real world. However, they rely on systems that are complex [4] and limited to a small range of haptic experiences [43].

2.4.2 Passive Haptics in Virtual Reality. In contrast, passive haptics uses physical objects of different materials and building techniques as props in VR, relying on the fact that physical props are a more feasible approach with natural feedback qualities [4, 39, 43]. Some authors have presented interesting approaches with single-purpose passive props. Yan et al. [74] created *VR Grabbers*, a passive chopsticks-like VR controller for precise virtual object manipulation that functions upon *ungrounded haptic retargeting* technique. Muender et al. [43] evaluated the effect of tangibles with different haptic fidelities (uniform-shaped objects, LEGO-built figures, and 3D-printed tangibles) on immersion, performance, and intuitive interaction for a 3D scene created in VR. Chang et al. [11] described *TASC*, a system of tangible objects for spatial puzzle solving.

Since it is impractical to map each virtual object to a physical proxy, authors have directed their work to developing multi-purpose or reconfigurable objects. In recent times, Arora et al. [4] presented *VirtualBricks*, a LEGO-based toolkit that enables construction of a variety of controllers and props for VR. This toolkit was shown to be highly versatile through a rich set of applications, including re-implementation of artifacts from past work [11, 39]. *HapTwist* [77] is a low-priced twistable passive device made of Rubik's Twist to create haptic proxies for distinct hand-graspable VR objects, such as ping-pong paddle, steering wheel, machine gun, and fishing rod. Also, Cheng et al. [12] developed *iTurk*, a foldable and reconfigurable object to represent a suitcase, a fuse cabinet, a railing, and a seat. Furthermore, Simeone et al. [59] explored the concept of Suzuki et al.'s *Substitutional Reality* [63] in VR by pairing pieces of physical furniture, with some degree of discrepancy, to their virtual counterparts.

2.4.3 Dynamic Passive Haptics. Zenner and Kruger [75] constructed *Shifty*, a weight-shifting mechanism to enhance object perception in VR. The device consisted of an array of mechanical actuators providing mixed active and passive haptic feedback. To reduce lack of generality of passive and active haptics devices, they introduced the concept of *dynamic passive haptic feedback*—systems that use actuators to change their passive haptic properties (size, shape, texture, weight, position, etc.) without exerting noticeable active forces on people. Other work includes shape and/or weight changing systems [39, 58, 62, 76], force-feedback systems for fingers and hands [14, 15, 24, 38, 48, 49].

This large body of work presents many possible opportunities to provide the feeling of manipulating physical objects in VR. Our work builds on this work by focusing on determining a vocabulary of how to leverage physical props to perform actions (i.e., gestures) in a VR system. While we think our work could help inform the design of gestures that incorporate active haptics, the scope of our work is currently limited to the use of props with passive haptics. Nonetheless, our intention is to provide both a gesture set that could be used by this other work, and to inform future elicitation studies that incorporate the use of physical props (or the feeling of them) when performing a gesture.

3 METHODOLOGY

We designed a user study to investigate the following research questions:

(1) How would people use physical props to manipulate virtual objects?
(2) What gestures would people perform with physical props to complete CAD-like and open-world referents in VR?
(3) What physical props would people choose to manipulate objects in VR?

Our study follows Wobbrock et al.'s elicitation method [71]. First, we present the effect of a referent in VR being completed. Second, we ask participants to choose one or more props from a group of 95 (75 props and 21 LEGO bricks; Figure 3) randomly arranged and numbered on a grid. Third, we ask participants to perform manipulative gestures with props that would complete the shown referent.

3.1 Selection of VR Referents

We aimed to select a rich and varied vocabulary of referents to ground our problem. We thus chose referents via domains (i.e., CAD and open-world games) rather than aiming for a specific quantity. "Rotate", "scale", and "zoom in", found in past gesture-based user interfaces for CAD 3D modelling [29, 34], were included. We expanded the vocabulary by adding other more-complex ones such as "bend", "bevel", "extrude", "colour", and "twist", typically found in AutoCAD or Blender. We also used games as a source for inspiration, because current VR applications on the market are overwhelmingly games, and including referents from that domain helps show that gestures with props can be used in other contexts. We picked "open/close a door", "turn on a light", and "darken a sky", extracted from RecRoom and Job Simulator. While we provide labels for our referents, these

Fig. 2. Gesture elicitation study with props that follows a previous methodology [71]. A participant: a) watches the completion of a referent, b) chooses a set of physical props, and c) and d) performs a gesture with the chosen prop(s) that would complete the referent.

were not shown to participants (they instead viewed a sample referent in a laptop). Our VR referents are shown alongside the proposed gestures in Figure 5 and Figure 6.

3.2 Selection of Physical Props

An underlying intention of elicitation studies is to design gestures/systems based on end-users' desires, rather than designers' intentions [42]. With gestures, the only limitations typically imposed are those of the system (e.g., a touch surface [71]) or human ability. For props, we mimicked the lack of designer imposition by having a large prop set, and wanted to make as few assumptions about what was "relevant" to participants. We developed a vocabulary of props that covered dimensions from sandtray therapy [25, 27], a form of therapy in which clients use *free association* [36] with physical objects arranged in categories (e.g., nature, tools, games, etc.) to create a narrative for a therapist in a tray of sand. While we had no interest in participants engaging in therapy, we were highly interested in allowing them to freely associate with the physical artifacts. We also took inspiration from prior research [4, 39, 43, 69] to include phone-related items, armory, avatars, tools, and office supplies.

We added more depth to our vocabulary by including (1) objects with different geometries (e.g., cube, plane, gears, tire), (2) objects with movable parts (e.g., mace, 3 DoF mechanism, glasses), (3) common household objects (e.g., fork, key, lock, knife, toys), and (4) VR and gaming controllers.

We gave our participants some variety in materials, and the option to use existing controllers. We used an *Eden260v* 3D printer to fabricate more than half of our props. 25 props were printed with rigid material, and 25 props were printed with flexible material, for a total of 50 distinct props fabricated with 3D printing technology. The next 8 props were LEGO-built assemblies (e.g., ghost, plant, car), and the following 21 were spare bricks that would allow our participants to construct a desired prop that was not available. We complemented the vocabulary with 13 retail-manufactured objects and 3 handheld controllers from Vive, Oculus and Nintendo Switch. Our vocabulary had 95 props in total as we did not want to unnecessarily constrain participants with a limited number of props nor to cognitively overwhelm them with a large number. Sandtray therapy suggests that 100 props are appropriate to avoid these extremes and inspire free association [36].

3.3 Participants

We recruited 21 participants, ages 19 to 32 ($M = 25.4$, $SD = 3.70$), 7 identified as female and 14 as male; each was paid $15. We assessed their skills and familiarity with VR and 3D printing using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) on two criteria. 6 participants reported never having used VR technologies. The remaining 15 reported they had used VR technology and rated themselves on the following: "I am skilled at using VR technology" ($M = 3.07$, $SD = 1.24$). 12 participants reported never having used 3D printing equipment. The remaining 9 reported they had operated 3D printers and rated themselves on the following: "I am skilled at operating 3D printers" ($M = 3.00$, $SD = 1.05$).

3.4 Apparatus

Participants watched the execution of the referents using a Vive headset connected to an MSI VR-ready laptop. We used a compact camera facing towards the participant's seat to record gestures, and a GoPro camera attached to the ceiling to capture the selection process of the props for gesture demonstration afterwards. The virtual objects were modelled in the Blender software and the animation of the referents was created in the Unity 3D game engine.

Fig. 3. Our vocabulary of physical props provided to participants in our elicitation study. Top: Grey- and white-coloured props were 3D printed with flexible and rigid material, respectively. We also include both LEGO-built specific figures, such as a flower or a fish, and spare bricks to allow free customization if a desired prop was not available. The vocabulary was complemented with retail-manufactured objects. Bottom: Number of gestures from Figures 5 and 6 for which each prop in the context-free grammar can be used for.

3.5 Procedure

Upon arrival, participants were compensated and asked to complete a consent form. They were then given a verbal description of the experiment. The VR referents were not revealed, but a sample was shown on a monitor. The props were randomly arranged on a table. The experimenter pointed out the different fidelities available to the participant: 3D-printed rigid and flexible, LEGO figures/bricks, and retail-manufactured objects. Participants were then given 5-10 minutes to familiarize themselves with the props. The experimenter then supervised and, upon request, supported participants for appropriate and comfortable wearing of the VR headset.

Our software randomly presented the 20 referents shown in Figure 5 and Figure 6. As in Wobbrock et al. [71], for each referent, participants performed a gesture with props. For each gesture, the experimenter asked each participant to indicate when the gesture was about to begin and when it was completed. No restrictions were imposed on the number of props to pick, and there was no time limit. The study concluded with a computer-administered demographics questionnaire.

4 RESULTS

Our results include a proposed participant-defined set of gestures using physical props, descriptions of patterns found in the elicited gestures, and two sets of agreement scores: one for gestures and one for physical props.

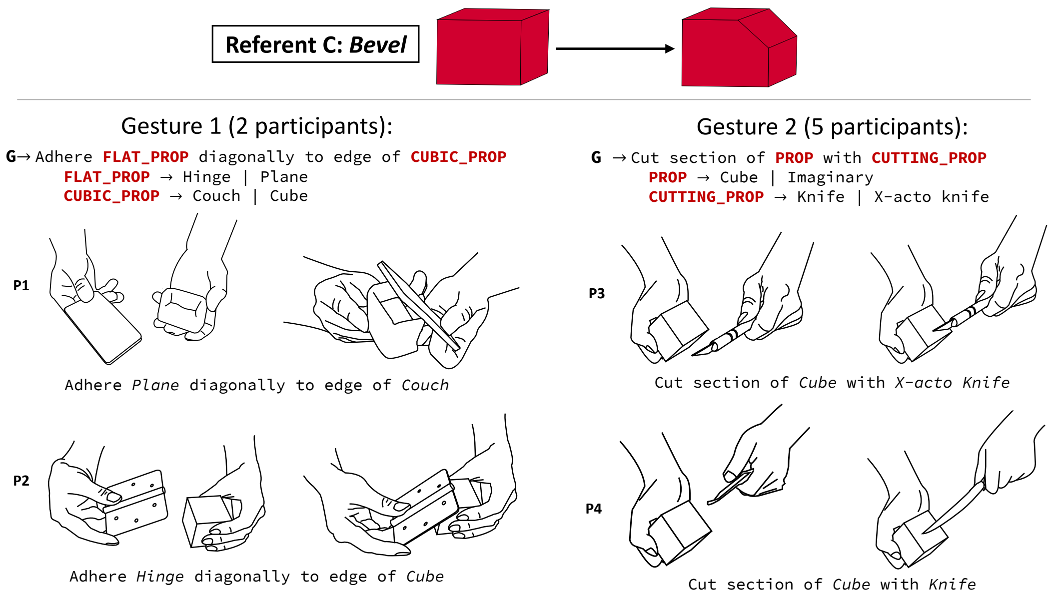


Fig. 4. Two of the thirteen gestures identified for referent C: *Bevel*. Each gesture is articulated as a context-free grammar, showing what physical props can be used in what role with that gesture.

4.1 Gesture Identification and Coding

We grouped participant responses into a set of gestures for each referent. For each referent, gestures that exhibited the same pattern were grouped together, and the group with the largest consensus was chosen to be the representative gesture for the referent. This differs slightly from elicitation studies for 2D systems (e.g., tabletops [71] and mobile platforms [23, 37, 54]), as we loosened the constraints from "gestures must be identical" to "gestures must be similar" or "gestures must show the same core pattern". We identified 182 unique gestures from the 420 gestures elicited from 21 participants for each of the 20 referents in our study.

One author qualitatively coded each gesture from the 21 participants. The author began with open coding [17] and reflective note-taking, where descriptive labels were assigned to the phenomena (gestures) observed during the time frames set by the participants.

Figure 4 illustrates this process for two gestures performed for referent C (bevel), where a rectangular prism's corner is beveled. The descriptive label for what P1 performed was "*associated the couch with the virtual model, and placed the plane diagonally on the arm rests of the couch*", and the description for what P2 performed was "*associated the cube with the virtual object and located an open hinge diagonally at one of the top edges*". The coder described what P14 performed (not shown in Figure 4) as "*associated the cube with the virtual object, held it with left hand, located their right palm diagonally at one of the top edges, and pushed inwards*". Despite involving different props, the gestures performed by P1 and P2 looked very similar (a flat prop placed diagonally on the edge of a cubic prop), so the coder considered P1 and P2 to have made the same gesture. Contrarily, prop agreement did not exist, as each used a different set. P14's gesture might sound very similar, but the inward "push" with the palm led the coder to consider gesture disagreement with P1 and P2. Nonetheless, there was partial prop agreement between P2 and P14, as both chose the cube.

In discussions with the other authors, the coding author merged these written descriptions (with "hard-coded" prop names) by including sets of the props allowed in each role. These were then labelled as variables to describe the role, resulting in a representation resembling a context-free grammar (CFG). As we continued analysis, we found that CFGs were a natural and expressive way of capturing how a given gesture could involve different props in similar roles, and of describing those roles. We grouped the two gesture variations of Gesture 1 in Figure 4 with *Adhere*. We grouped the couch and the cube in the variable "CUBIC_PROP", and grouped the plane and hinge in the variable "FLAT_PROP". CUBIC_PROP represents the virtual prism being beveled and FLAT_PROP represents the one causing the bevel effect:

```
G -> Adhere FLAT_PROP diagonally to edge of CUBIC_PROP
    FLAT_PROP -> Hinge | Plane
    CUBIC_PROP -> Cube | Couch
```

The most used gesture (five participants) for referent C (Bevel), however, was *Cut*, where participants took a cube, and then cut a bevel into it with either a knife or an X-acto knife. Here, we represent the virtual cube as the variable "PROP", which represents the object being beveled, and represent the knife and X-acto knife as "CUTTING_PROP", which affords the action of cutting:

```
G -> Cut section of PROP with CUTTING_PROP
    PROP -> Cube | Imaginary
    CUTTING_PROP -> Knife | X-acto Knife
```

Participants doing the *Cut* gesture either held a cube to represent the virtual cube, or held their hand as if they were holding an imaginary cube, then cut the bevel with the CUTTING_PROP.
For each referent (here, Bevel), we chose the most demonstrated gesture (here, *Cut*) for our proposed gesture set. Figure 5 and Figure 6 show our proposed gestures with their corresponding referents.

4.2 Gesture Agreement

In previous studies, participants suggested one gesture per referent, so agreement was calculated for gestures alone. In our study, participants offered both a gesture and a selection of props, and may have based their actions on either a gesture or a set of props. We thus analyze both gesture agreement and prop selection.

For gesture agreement, we use the formula presented by Vatavu and Wobbrock [68], given as:

$$AR_{gesture}(r) = \frac{\sum_{P_i \subseteq P} \frac{1}{2}|P_i|(|P_i|-1)}{\frac{1}{2}|P|(|P|-1)} \tag{1}$$

This formula is derived from pairwise comparison of participant gestures. For a given referent, we calculate the agreement by looking at each pair of participants, saying that the two participants agree if they picked the same gesture (similarity of gestures is 1), and do not agree if they picked different gestures (similarity of gestures is 0).

Figure 7a illustrates pairwise gesture agreement for the four participant *Bevel* examples introduced in Figure 4. Pairwise agreement can be represented as a complete graph, where nodes represent participants, and edge weights represent the similarity between those participants' gestures, i.e., either 0 (if they are different) or 1 (if they are the same). The agreement for a given referent is the sum of the weights of those edges divided by the total number of edges: $\frac{1}{2}n(n-1)$, where n is the number of participants. This score is thus bounded between 0 and 1, as if everyone agreed and all edges were equal to 1, then the equation would be the number of edges divided by the number of edges. While previous formalization of agreement scores [68] did not represent agreement as a graph, we find this representation useful when reasoning about prop agreement.

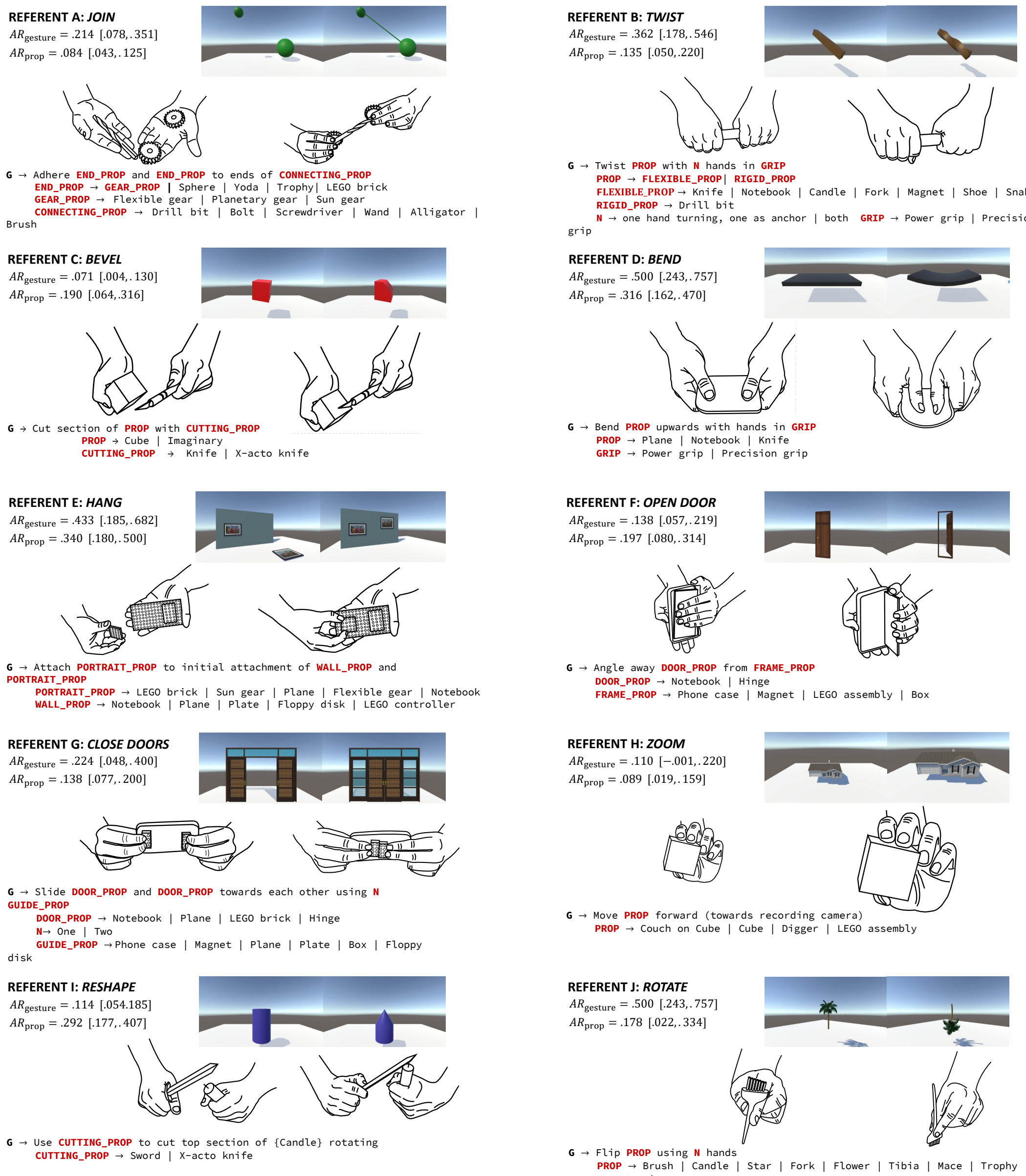


Fig. 5. Gesture set with 3D physical props for referents A to J. For each referent, the preferred set of props and the preferred gesture are presented; the initial state of the most common prop-gesture interaction is shown on the left side, and the end state is shown on the right. We calculate gesture and prop agreement scores using Equation 1 and Equation 2, respectively. We also include 95% confidence intervals from jackknifing [52] as recommended by Tsandilas [65].

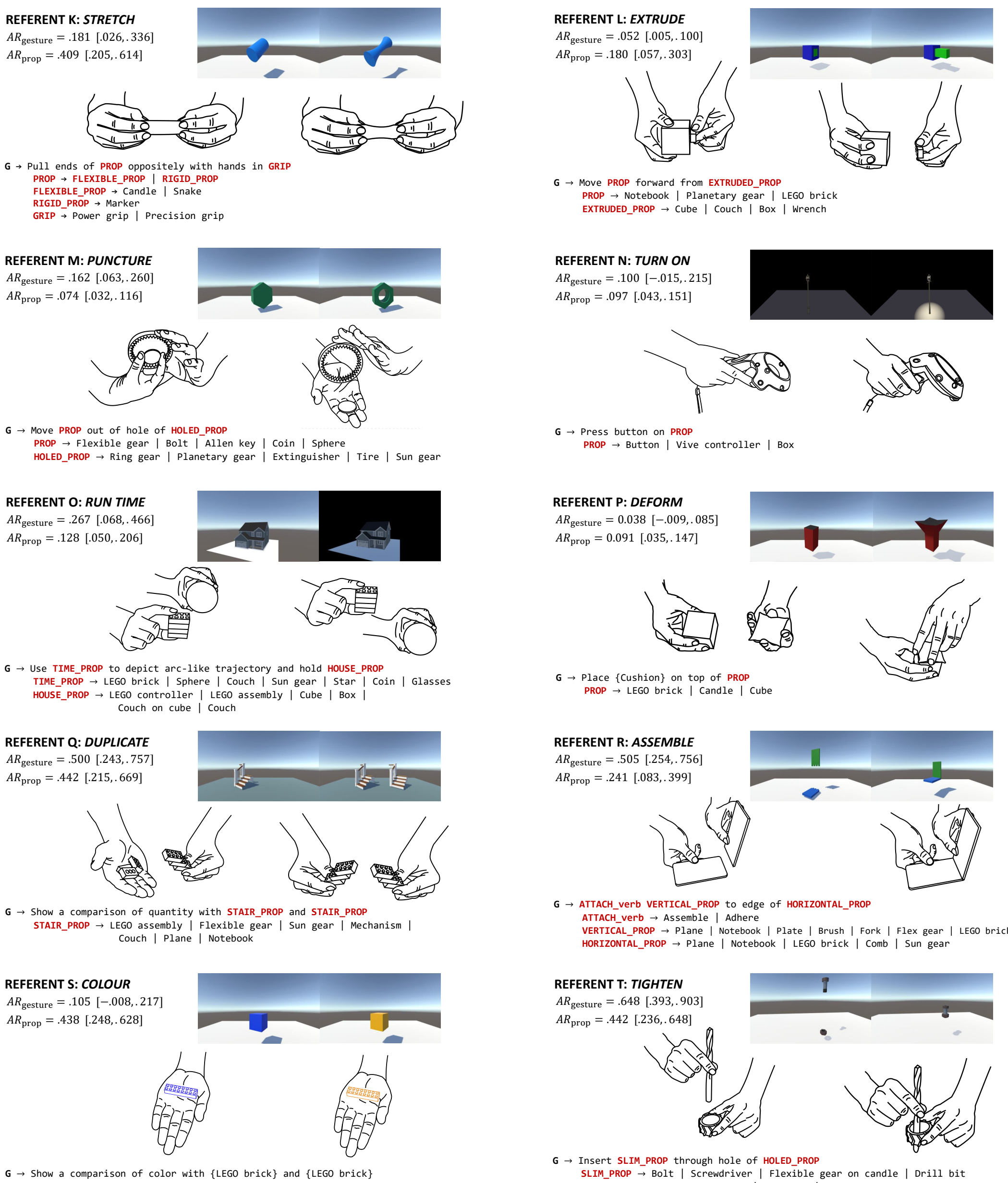
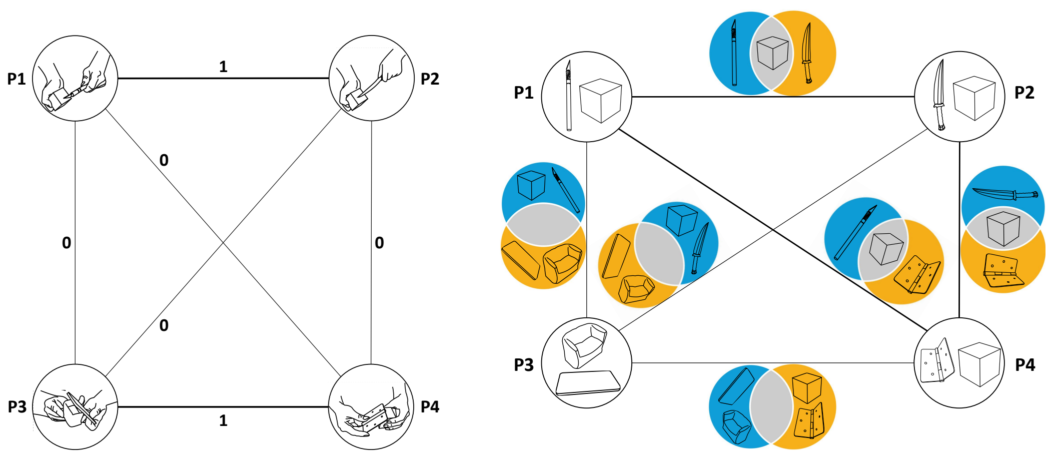


Fig. 6. Gesture set with 3D physical props for referents K to T. For each referent, the preferred set of props and the preferred gesture are presented; the initial state of the most common prop-gesture interaction is shown on the left side, and the end state is shown on the right. We calculate gesture and prop agreement scores using Equation 1 and Equation 2, respectively. We also include 95% confidence intervals from jackknifing [52] as recommended by Tsandilas [65].



(a) Similarity graph for gestures, demonstrating either full agreement (same gesture, 1) or not (different, 0).

(b) Similarity graph for props, demonstrating partial agreement depending on the number of overlapping props.

Fig. 7. Complete graphs that capture gestures with physical props performed by 4 participants for referent C: *Bevel*. For gestures, similarity between two sets is the weight of the edge that joins them. For props, we measure similarity of props sets, which is the weight of the edges that joins them. In both cases, we calculate agreement by summing the similarity of each pair of sets compared, and dividing by the total number of pairs that could be identical.

4.3 Prop Agreement

In our study, participants selected props in addition to gestures. We were interested in seeing how much prop selection agreed across participants for a given referent. However, prop agreement is more complicated than gesture agreement: participants selected *sets* of physical props, which means that there can be partial agreement between participants.

To analyze prop agreement, we propose a generalization of the Vatavu and Wobbrock's agreement rate formula [68]. Instead of using simple equality between pairs, we use a different measurements of set similarity, here denoted by the function SIM:

$$AR_{prop}(r) = \frac{\sum_{i=1}^{P-1} \sum_{j=i+1}^{P} SIM(P_i, P_j)}{\frac{1}{2}|P|(|P|-1)} \tag{2}$$

We define Equation 2 as the sum of the similarity of each pair of sets by the total number of pairs of sets that could be identical, where P is the total number of people. P_i and P_j group all possible pairs of sets for r. Equation 2 is structurally similar to the simple gesture agreement formula (Equation 1), but this gives us more flexibility in comparing participants' choices. We can thus construct a connected graph to represent prop selection similarity, where nodes represent participants, and edge weights represent the similarity between those participants' gestures (a real number between 0 and 1 inclusive).

In the graph representation (Figure 7b), each edge is weighted by SIM, the set similarity between their sets of selected props, rather than just 1 (same) or 0 (different). If SIM is bounded between 0 and 1, then the overall agreement (sum of the weights of all edges divided by the number of edges)

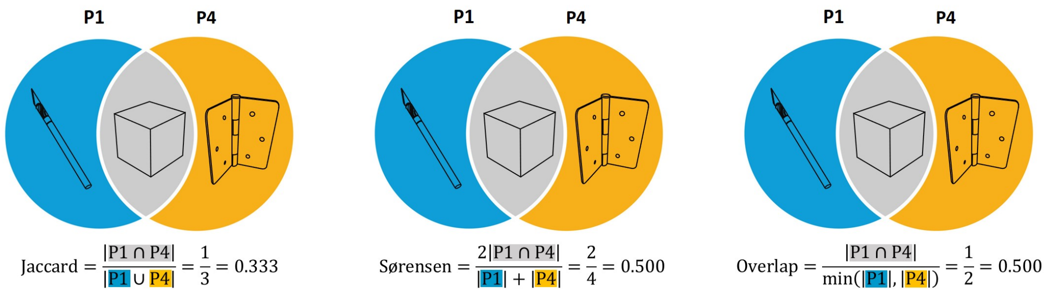


Fig. 8. Venn diagrams showing a pairwise comparison between prop sets of P1 and P4. We propose three measures of set similarity to evaluate partial agreement: Overlap, Sørensen, and Jaccard. Overlap is the most optimistic, leading to the highest agreement. Jaccard is the most pessimistic, leading to the lowest agreement. Sørensen prevails between the other two (or equal to one of them). In this pairwise comparison, Sørensen and Overlap are the same.

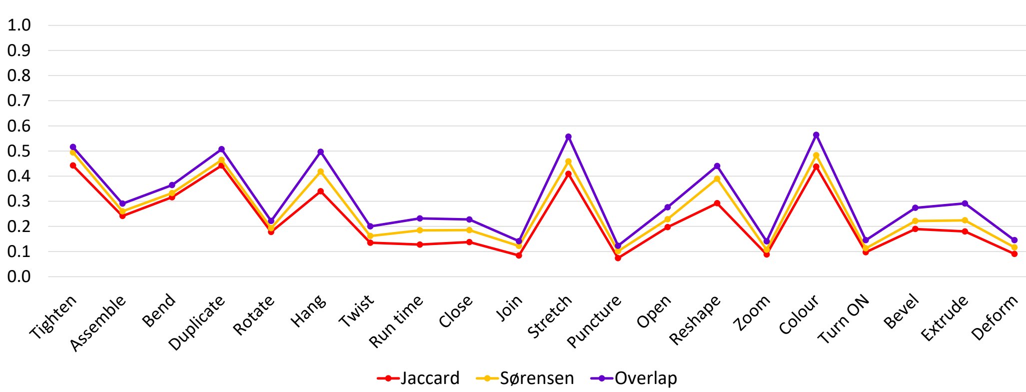


Fig. 9. Prop agreement scores for each VR referent with the three different similarity metrics. Taking an Overlap approach leads to the highest agreement, whereas going with Jaccard leads to the lowest agreement. Sørensen-Dice is between the two.

is also bounded between 0 and 1. To be analogous to gesture similarity, SIM must also be 0 if there are no props in common, or 1 if they are completely equal (analogous to choosing different gestures or the same gesture).

Figure 8 shows the three set similarity metrics we have identified for SIM: the Jaccard Index ($Jaccard$), the Overlap coefficient ($Overlap$) and the Sørensen-Dice coefficient (*Sørensen*). These three metrics can all satisfy our requirements for a similarity score, and if they are sets of size one (e.g., a single gesture per participant), they collapse into a binary 1 for equality and 0 for inequality. In other words, all three metrics are equivalent to Vatavu and Wobbrock's score when comparing a single agreement.

Figure 9 shows the prop agreement score for all three metrics (Jaccard, Sørensen, Overlap) for each referent. The three metrics give similar results for each referent. Jaccard is consistently the lowest agreement, Overlap is the highest, and Sørensen is in-between the two. In our analysis, we use the Jaccard Index as the the most conservative metric for identifying high agreement.

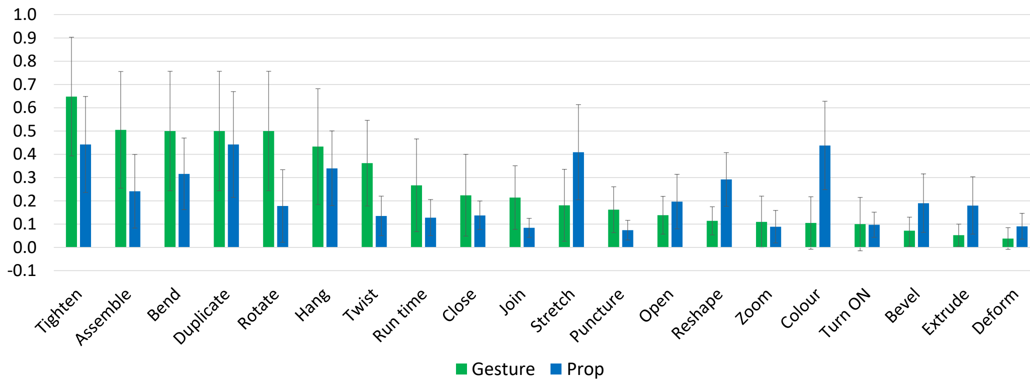


Fig. 10. The two agreement scores (using the Jaccard index) sorted in descending order of the gesture agreement score. Error bars represent 95% confidence intervals from jackknifing [52].

Finally, Figure 10 reports our gesture and prop agreement (Jaccard Index) scores for each referent, sorted by gesture agreement score. For some referents, the two scores agree; for others, the gesture agreement is higher, and for others, the prop is lower.

Elicitation studies can introduce bias [42, 55, 65], and bias can falsely inflate agreement scores by increasing the likelihood of chance agreement [65]. We thus calculated overall chance agreement and chance-corrected agreement scores for gestures using Scott's [57] π and Fleiss' [21] κ_F, to increase transparency and identify agreement occurred by chance [65] in our study. To interpret agreement, we also calculated confidence intervals of both gesture and prop agreement scores for each referent using the *Jackknifing* [52] re-sampling method. We avoided using the agreement levels suggested by Vatavu and Wobbrock [68] to interpret agreement, as these are flawed and more research is needed in this area [65]. Our overall gesture agreement score was .261 and our chance agreement was .016, meaning that our chance-corrected agreement score yielded .249. Our 95% confidence intervals of both gesture and prop agreement for each referent can be observed in Figure 5, Figure 6, and Figure 10. In section 6, we describe the rationale behind not reporting chance agreement for props as more research is needed on this matter.

5 DISCUSSION

Our work has implications for VR systems that use physical props and future elicitation studies.

5.1 Props were selected for shape and affordances

When selecting props, our participants chose objects that resembled the objects in the virtual world and had affordances [22, 46] for the actions they wanted to take. For example, in *Bevel*, participants chose a cube as a stand-in for the cube in the virtual world, and a Knife or an X-acto Knife because both afforded cutting the corner away to create the bevel. In referents B (*Twist*) and D (*Bend*), participants picked a single prop that resembled the shape of the virtual object and that was flexible to be twisted or bent; one exception was the drill bit, a rigid object that resembled the action in its spiral shape.

In some cases, the affordances chosen were metaphorical. In referent N, *Turn On*, participants chose objects that supported pressing (i.e., as a button) to turn on the light. Because objects were chosen for their affordances, and each object had multiple affordances (from its shape, meaning, structure, or material), objects could serve different purposes in different gestures: the knife, for example, afforded cutting in *Bevel* and *Reshape*, but bending in *Twist* and *Bend*.

5.2 A small set of props can handle many gestures

A small number of props covered a large number of gestures. As seen in Figure 3, only 48 of the 95 props were chosen; 18 of these were only used for a single gesture. In addition, most prop variables had on average 4.2 props that were selected to fulfill a certain use. This means that a small set of props could support a large set of gestures. In fact, 12 props and a group of spare LEGO bricks would be sufficient to handle all 20 gestures: sun gear, flexible gear, x-acto knife, cube, candle, phone case, notebook, plane, Vive controller, tire, drill bit, and cushion would enable all gestures.

5.3 Props for gestures vs. gestures for props

We presented two separate ways of measuring agreement, which correspond to different strategies by participants: choosing their gesture first and finding props to support that gesture, or picking a desired prop and then figuring out a gesture to perform. As such, we offered two different agreement scores to indicate how consistent people were in their choices.

In some cases, the gestures were likely chosen to fit the props. In Figure 10, referents like *Bevel*, *Colour*, *Reshape*, and *Stretch* exhibit a high prop agreement score, denoting that only a few props picked for those referents were used in a variety of gestures, such as the cube or the candle.

In other cases, props were likely chosen to fit the gesture. Referents like *Twist*, *Rotate*, and *Assemble* show a high gesture agreement score compared to the prop score, thus suggesting that a lot of props afforded the most used gestures. In these cases, it may be more important to recognize the gesture than the props that were chosen in an implemented system.

We iterated through several attempts at a combined prop-gesture agreement score, but ultimately decided it was outside the scope of this paper. First, developing such a score would need to consider the structure of how props were chosen for different gestures, possibly categorizing the objects by their affordances. Second, it could be that the agreement score should depend more on the prop choice in some cases and more on the gesture in others (e.g., through weights), depending on the referent; more analysis would help understand when each choice was important.

5.4 CFGs allow for expressive and programmable gestures but do not account for performance

When grouping gestures, we found a natural way to express them was using context-free grammars (CFGs). This notation captures the variations in props that each gesture had, and groups together the types of props that were used in a certain role in that gesture. For instance, with *Close*, a notebook, a plane, two LEGO bricks and a hinge in its flat position could be used as “DOOR_PROP”; presumably other objects could be used to cut as well.

Using a formalism like a CFG also enables system designers to efficiently describe gestures for a sensing system; if other props not used in our study could fulfill a role, then they could simply be added to the list. Combined with a method for detecting objects and recognizing grammars, then designers could adapt and define gestures to various situations.

Our CFGs, however, do not compile performance information, such as magnitude, force, or speed and would need to be adapted or expanded to do so. For example, for referent E (hang), they could indicate where on `WALL_PROP` the `PORTRAIT_PROP` is located, or for referents F and J (open door and rotate), they could note the number of degrees. For referent K (stretch), an adapted or expanded CFG could inform how much a prop is stretched.

5.5 Our new agreement score enables new elicitation studies

To measure agreement between participants’ prop selections, we generalized participant agreement scores to handle sets of selections. When participants make a single selection, our generalized agreement scores are equivalent to the agreement score formula first presented in [71] and later formalized in [68]. However, with this new agreement score, researchers and designers can study agreement in new situations.

For example, multimodal input methods, such as gesture and speech input, previously had to look at agreement separately [34]. If participants are able to choose either a gesture, a voiced utterance, or both, then our generalized agreement score could capture partial overlap between what participants suggest, rather than studying modalities independently. Any of the three similarity metrics we suggested (Jaccard, Sørensen, Overlap) will give a 1 for full equivalence, a 0 for no overlap, and an intermediate value for partial overlap.

Our generalized agreement score can also handle elicitation of sets from participants. If a researcher wanted to follow up with a study on preferred props for a referent, they can use our measure to identify which referents have an agreement score. Alternatively, this could work for other modalities: suggesting multiple gestures or multiple commands as options. This technique could also be used for finer-grained comparison of gestures: if two participants provide a gesture for a referent, and they use different sets of fingers, a partial agreement score could capture similarities between gestures.

5.6 Gestures were typically used for one referent

Our chance-corrected agreement score was only marginally different from our overall gesture agreement score calculated with Equation 1. This result indicates that gesture reuse [65] did not happen often in the study.

Generally, gestures proposed for a given referent were rarely proposed for another one. "`Fold hinge with hands`" and "`Shake wand with hand`" were elicited for up to three referents maximum, but none of these gestures were the representative gesture for any referent in our gesture-prop set.

5.7 Implications for design of haptic systems for VR

To employ our findings, designers of future VR systems need to reflect on practical considerations for the props sets that they choose: prop and gesture recognition, space limitations, and possibly prop reusability.

To recognize props and gestures suggested by our findings, researchers and designers can leverage existing sensing methods. For example, gestures for *open door*, *rotate*, *twist*, *bend*, and *stretch* involve clearly defined and timed rotations, translations, and deformations, which can be sensed via electromechanical sensors, but gestures for *colour*, *bevel*, and *join* might better use vision-based systems. The specific sensor could depend on accuracy needs.

Designers should also consider physical space challenges, as there may be a need for a dedicated space to place props [77] that the user can easily move and pick up props from. People can recognize and manipulate objects by touch alone without looking at them (e.g., keys, phone). Thus, users could carry props on their bodies (e.g., in a type of pocket or tool-belts). Alternatively, props could be fixed and available at all times using approaches like *Haptic-go-round*'s [28]. We are agnostic about drops or fumbles, but multiple possible solutions are available, such as tracking, tethers on props, passthrough cameras, or even AR instead of VR.

Furthermore, we reported that props were selected for shape and affordances. Researchers, for instance, may consider designing multi-purpose, reconfigurable, or reusable props that incorporate these shapes and affordances. Reusability reduces the need and the cost of creating multiple haptic props [77]. Our results show that 12 props and a group of LEGO bricks are enough to handle the user-defined gesture set presented in Figure 5 and Figure 6. Versions of *HaptoBend* [39] and *HapTwist* [77] could incorporate some of the different shapes of our 12 props, and *VirtualBricks* [4] could offer the group of LEGO bricks as well as shapes.

To illustrate how a VR system using this work could look, consider the following scenario:

> Deganawida, an architecture student, wants to edit and evaluate the interior of a virtual house. They use a VR CAD application to get an intuitive sense of the size of the house, and they carry the props in a tool-belt. Deganawida begins by adding a wall between the living room and the kitchen, for which they pull out a cube rendered with Haptobend [39] and a LEGO brick from the belt to do an extrusion (Referent L: extrude). A proximity sensor on the cube senses this gesture (it senses when the brick is moved away). After this, they reconfigure Haptobend [39] as a plane and use it with the brick to hang paintings and shelves on the wall (Referent E: Hang). This gesture is recognized by a vision-based system. Deganawida moves the Haptobend object to specify a coarse region on the wall (highlighted in a rectangle), and the position of the brick to provide finer control of the painting location. Deganawida then returns both props to the belt, and pulls out the tire and the sun gear to make a circular window (Referent M: Puncture) on another wall, controlling the position in the same way. Another proximity sensor detects when the sun gear has been taken out and away from the center of the tire. Finally, Deganawida wishes to see the interior of the house at night time. To accomplish this, they exchange the tire by the Haptobend configured as a cube and depict an arc around it with the gear (Referent O: run time). This gesture is sensed by a built-in accelerometer in the gear. After this, the architecture student has completed their edits, and walks around the space to take it all in.

In this scenario, having user-defined gestures and props could mean higher immersion, or easier or more memorable inputs. Deganawida wants to feel immersed in the house, so increasing presence could lead to a better experience and better results. Meanwhile, having access to props that can represent different objects in the environment might enable more learnable or expressive commands to accomplish the tasks they would like to accomplish.

While future work is needed to test the effect and implementation of gestures with props, our findings give an initial specification of what gestures to support and how to determine them.

5.8 Our user-defined gestures fit partially in existing gestures taxonomies

Our gesture set fits partially in Karam and schraefel m. c. 's [32] taxonomy of gesture-based human computer interactions. For the *gesture style* category, our gestures fall under the *manipulative gestures* subcategory, as they are interactions with tangible objects used to represent digital objects. With regard to the *application domain*, our work belongs to the *Virtual and Augmented Reality* subcategory. At the current stage of our research, we are agnostic about *enabling technology* and *system response*.

Choi et al.'s [13] taxonomy seems insufficient for our gesture set, as it does not include props. Without considering props, our gestures in Figure 5 and Figure 6 could be classified within this taxonomy. 16 of the 20 are *two hands* gestures, and the remaining 4 are *one hand* gestures. Gestures for all referents, except referents Q (duplicate) and S (colour), can be considered *dynamic*, as there is a *hand transition* and/or a *hand shape transition* while being performed.

However, it is important to note that these existing taxonomies were insufficient for coding of gestures in our study, as they did not adequately capture the prop+gesture division that is at the heart of our current exploration, and so we opted for an open-coding technique.

6 LIMITATIONS AND FUTURE WORK

A limitation of our study is that our participants were undergraduate and graduate students from different disciplines. Professional designers and architects might behave differently. We focus on a set of manipulative gestures usable by non-technical people to do CAD 3D modelling in VR or perform in an open world game, enhancing immersion, intuitiveness, and realism with the incorporation of physical props. Another limitation of this work is that our gesture-prop set requires validation. A follow-up study may require showing video footage of the gestures to a new group of participants to investigate if they can infer the intended referent and to whether the gestures are usable and memorable.

Another limitation of our study is that participants removed the VR headset to perform their gestures with props. This choice was an intentional part of our study design, as we did not want to limit the actions that were possible with physical props, and is similar to previous elicitation studies that do not provide system feedback to participant action [3, 9, 35, 50, 54, 71]. Future work could introduce props to VR using one or more of the methods discussed in subsection 5.7, and a similar study design could elucidate gestures with props that work entirely in VR. Nonetheless, the current work can still be used to guide the design of systems that incorporate props in VR and our results are not limited to a single design solution.

While we asked our participants to think aloud as they gestured, as recommended by Wobbrock et al. [71], our participants did not always engage with this practice and the experimenter did not remind participants to keep talking. More verbalization from participants would help us to better understand (1) what attributes and affordances they look for in props and (2) what their gestures intend to represent. Future work might consider methods such as retrospective think-aloud or a more rigorous adherence to think-aloud protocols [47].

Also, chance agreement and chance-corrected agreement for *props* remains unknown. While Tsandilas [65] shows how to determine chance agreement for gestures in elicitation studies, determining chance agreement for props is more complicated. Further research is needed to explore how Tsandilas' recommended metrics could be applied for props. If participants were limited to use one prop per referent, they might be suitable. However, participants form props sets, meaning that both the size of the vocabulary of props and the chance of choosing one prop over another vary on-the-fly. Moreover, does the number of virtual objects in a referent bias participants' selection criteria? Should researchers explore chance agreement of shapes and affordances, instead of the actual props? Future work should explore these questions to estimate chance agreement for props.

We also plan to study the relationship between gesture choice and prop choice, formalize the affordances for elicited props, and explore how gestures with physical props could be used in a working system, such as a VR CAD environments or open-world video games.

7 CONCLUSION

In this paper, we have presented a VR elicitation study leading to a user-defined set of gestures with physical props for 20 CAD-like and open-world game referents based on participants' multimodal agreement. To do this, we generalized the previous unimodal agreement metric to account for multiple selections (gesture + prop), which we believe will help researchers design multimodal gesture sets with more precision, versatility, and expressivity. We also present a context-free grammar to describe mid-air manipulations of props as well as a classification of gestures with props. In eliciting gestures with physical props for VR referents, we have gained insight into mental models of non-technical people and translated them into implications for a VR system that works upon natural and intuitive manipulation of physical props.

ACKNOWLEDGMENTS

Thank you to the University of Waterloo's Touchlab, Haptic Experience Lab, and the Games Institute. This work was made possible by NSERC Discovery Grant 2016-04422, NSERC Discovery Grant 2019-06589, NSERC Discovery Accelerator Grant 492970-2016, NSERC CREATE Saskatchewan-Waterloo Games User Research (SWaGUR) Grant 479724-2016, and Ontario Early Researcher Award ER15-11-184.

REFERENCES

[1] Parastoo Abtahi and Sean Follmer. 2018. Visuo-Haptic Illusions for Improving the Perceived Performance of Shape Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Volume 2018-April. ACM Press, New York, New York, USA, 1–13. http://dl.acm.org/citation.cfm?doid=3173574.3173724.

[2] Sema Alaçam. 2014. The Many Functions of Hand Gestures While Communicating Spatial Ideas An Empirical Case Study. In *Proceedings of the XVIII Conference of the Iberoamerican Society of Digital Graphics - SIGraDi: Design in Freedom* number 2006. Editora Edgard Blücher, São Paulo, (December 2014), 106–109. http://www.proceedings.blucher.com.br/article-details/14234.

[3] Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, (May 2016), 3822–3833. https://dl.acm.org/doi/10.1145/2858036.2858385.

[4] Jatin Arora, Aryan Saini, Nirmita Mehra, Varnit Jain, Shwetank Shrey, and Aman Parnami. 2019. VirtualBricks: Exploring a Scalable, Modular Toolkit for Enabling Physical Manipulation in VR. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* number Chi. ACM Press, New York, New York, USA, 1–12. http://dl.acm.org/citation.cfm?doid=3290605.3300286.

[5] Autodesk Help. 2021. About create vr. Accessed May 31, 2021. Autodesk. Retrieved 05/31/2021 from https://knowledge.autodesk.com/support/alias-products/learn-explore/caas/CloudHelp/cloudhelp/2020/ENU/Alias-CreateVR/files/Alias-CreateVR-aboutcreatevr-html-html.html.

[6] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces - ISS '16*. ACM Press, New York, New York, USA, (May 2016), 501–504. http://dl.acm.org/citation.cfm?doid=2992154.2996883.

[7] Gilles Bailly, Thomas Pietrzak, Jonathan Deber, and Daniel J. Wigdor. 2013. Métamorphe: Augmenting Hotkey Usage with Actuated Keys Gilles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, New York, New York, USA, 563. http://dl.acm.org/citation.cfm?doid=2470654.2470734.

[8] Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. 2016. Normaltouch and texturetouch: high-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (UIST '16). Association for Computing Machinery, Tokyo, Japan, 717–728. https://doi.org/10.1145/2984511.2984526.

[9] Sarah Buchanan, Bourke Floyd, Will Holderness, and Joseph J. LaViola. 2013. Towards user-defined multi-touch gestures for 3D objects. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces - ITS '13*. ACM Press, New York, New York, USA, 231–240. http://dl.acm.org/citation.cfm?doid=2512349.2512825.

[10] Jeff Butterworth, Andrew Davidson, Stephen Hench, and Marc. T. Olano. 1992. 3DM: A Three Dimensional Modeler Using a Head-Mounted Display. In *Proceedings of the 1992 symposium on Interactive 3D graphics - SI3D '92*. Volume Part F1296. ACM Press, New York, New York, USA, 135–138. http://portal.acm.org/citation.cfm?doid=147156.147182.

[11] Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2017. Tasc: combining virtual reality with tangible and embodied interactions to support spatial cognition. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17). Association for Computing Machinery, Edinburgh, United Kingdom, 1239–1251. https://doi.org/10.1145/3064663.3064675.

[12] Lung-Pan Cheng, Li Chang, Sebastian Marwecki, and Patrick Baudisch. 2018. iTurk: Turning Passive Haptics into Active Haptics by Making Users Reconfigure Props in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Volume 2018-April. ACM Press, New York, New York, USA, 1–10. http://dl.acm.org/citation.cfm?doid=3173574.3173663.

[13] Eunjung Choi, Heejin Kim, and Min K. Chung. 2014. A taxonomy and notation method for three-dimensional hand gestures. *International Journal of Industrial Ergonomics*, 44, 1, (January 2014), 171–188. https://linkinghub.elsevier.com/retrieve/pii/S0169814113001285.

[14] Inrak Choi, Heather Culbertson, Mark R. Miller, Alex Olwal, and Sean Follmer. 2017. Grabity. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, (October 2017), 119–130. https://dl.acm.org/doi/10.1145/3126594.3126599.

[15] Inrak Choi and Sean Follmer. 2016. Wolverine: a wearable haptic interface for grasping in vr. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (UIST '16 Adjunct). Association for Computing Machinery, Tokyo, Japan, 117–119. https://doi.org/10.1145/2984751.2985725.

[16] Inrak Choi, Eyal Ofek, Hrvoje Benko, Mike Sinclair, and Christian Holz. 2018. Claw: a multifunctional handheld haptic controller for grasping, touching, and triggering in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3174228.

[17] John W Creswell and Cheryl N Poth. 2018. *Qualitative inquiry and research design: Choosing among five approaches*. (Fourth edition). Sage Publications, Thousand Oaks, CA, USA.

[18] Bruno R. De Araújo, Géry Casiez, Joaquim A. Jorge, and Martin Hachet. 2013. Mockup Builder: Direct 3D Modeling On and Above the Surface in a Continuous Interaction Space. *Computers & Graphics*, 37, 3, (May 2013), 165–178. https://linkinghub.elsevier.com/retrieve/pii/S0097849312001811.

[19] Alexandra Delazio, Ken Nakagaki, Roberta L. Klatzky, Scott E. Hudson, Jill Fain Lehman, and Alanson P. Sample. 2018. Force Jacket: Pneumatically-Actuated Jacket for Embodied Haptic Experiences. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–12. http://dl.acm.org/citation.cfm?doid=3173574.3173894.

[20] Seth M. Feeman, Landon B. Wright, and John L. Salmon. 2018. Exploration and evaluation of CAD modeling in virtual reality. *Computer-Aided Design and Applications*, 15, 6, (November 2018), 892–904. http://www.cad-journal.net/files/vol_15/Vol15No6.html.

[21] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 5, 378–382. http://content.apa.org/journals/bul/76/5/378.

[22] 1986. *The ecological approach to visual perception*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA. Chapter Eight: The Theory of Affordances, 127–136.

[23] Antonio Gomes, Lahiru Lakmal Priyadarshana, Aaron Visser, Juan Pablo Carrascal, and Roel Vertegaal. 2018. MagicScroll: A Rollable Display Device with Flexible Screen Real Estate and Gestural Input. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, (September 2018), 1–11. https://dl.acm.org/doi/10.1145/3229434.3229442.

[24] Xiaochi Gu, Yifei Zhang, Weize Sun, Yuanzhe Bian, Dao Zhou, and Per Ola Kristensson. 2016. Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, (May 2016), 1991–1995. https://dl.acm.org/doi/10.1145/2858036.2858487.

[25] Mark Hancock, Thomas ten Cate, Sheelagh Carpendale, and Tobias Isenberg. 2010. Supporting Sandtray Therapy on an Interactive Tabletop. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, New York, New York, USA, 2133. http://portal.acm.org/citation.cfm?doid=1753326.1753651.

[26] H. G. Hoffman, A. Hollander, K. Schroder, S. Rousseau, and T. Furness. 1998. Physically Touching and Tasting Virtual Objects Enhances the Realism of Virtual Experiences. *Virtual Reality*, 3, 4, (December 1998), 226–234. http://link.springer.com/10.1007/BF01408703.

[27] Linda E Homeyer and Daniel S Sweeney. 2016. *Sandtray therapy: A practical manual*. Routledge, New York, NY, USA.

[28] Hsin-yu Huang, Chih-wei Ning, Po-Yao Wang, Jen-Hao Cheng, and Lung-Pan Cheng. 2020. Haptic-go-round: A Surrounding Platform for Encounter-type Haptics in Virtual Reality Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, (April 2020), 1–10. https://dl.acm.org/doi/10.1145/3313831.3376476.

[29] Jinmiao Huang and Rahul Rai. 2014. Hand gesture based intuitive CAD interface. In *Proceedings of the ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 1A: 34th Computers and Information in Engineering Conference* number June. American Society of Mechanical Engineers, Buffalo, NY, USA, (August 2014), 8 pages. https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings/IDETC-CIE2014/46285/Buffalo,%20New%20York,%20USA/256676.

[30] Bret Jackson and Daniel F. Keefe. 2016. Lift-Off: Using Reference Imagery and Freehand Sketching to Create 3D Models in VR. *IEEE Transactions on Visualization and Computer Graphics*, 22, 4, (April 2016), 1442–1451. https://ieeexplore.ieee.org/document/7383322/.

[31] Hessam Jahani and Manolya Kavakli. 2018. Exploring a user-defined gesture vocabulary for descriptive mid-air interactions. *Cognition, Technology & Work*, 20, 1, (February 2018), 11–22. http://link.springer.com/10.1007/s10111-017-0444-0.

[32] Maria Karam and m. c. schraefel m. c. 2005. A Taxonomy of Gestures in Human Computer Interaction. Technical report 261149. University of Southampton, 45 pages. http://eprints.soton.ac.uk/id/eprint/261149.

[33] D.F. Keefe, R.C. Zeleznik, and D.H. Laidlaw. 2007. Drawing on Air: Input Techniques for Controlled 3D Line Illustration. *IEEE Transactions on Visualization and Computer Graphics*, 13, 5, (September 2007), 1067–1081. http://ieeexplore.ieee.org/document/4135646/.

[34] Sumbul Khan and Bige Tunçer. 2019. Automation in Construction Gesture and speech elicitation for 3D CAD modeling in conceptual design. *Automation in Construction*, 106, July 2017, 102847. https://doi.org/10.1016/j.autcon.2019.102847.

[35] Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. 2010. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '10). Association for Computing Machinery, Lisbon, Portugal, 239–248. https://doi.org/10.1145/1851600.1851640.

[36] Martin Kuška, Radek Trnka, Aleš A. Kuběna, and Jiří Růžička. 2016. Free Associations Mirroring Self- and World-Related Concepts: Implications for Personal Construct Theory, Psycholinguistics and Philosophical Psychology. *Frontiers in Psychology*, 7, June, (June 2016), 1–13. http://journal.frontiersin.org/Article/10.3389/fpsyg.2016.00981/abstract.

[37] Byron Lahey, Audrey Girouard, Winslow Burleson, and Roel Vertegaal. 2011. PaperPhone: Understanding the Use of Bend Gestures in Mobile Devices with Flexible Electronic Paper Displays. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, New York, New York, USA, 1303. http://dl.acm.org/citation.cfm?doid=1978942.1979136.

[38] Jaeyeon Lee, Mike Sinclair, Mar Gonzalez-Franco, Eyal Ofek, and Christian Holz. 2019. TORC: A Virtual Reality Controller for In-Hand High-Dexterity Finger Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–13. http://dl.acm.org/citation.cfm?doid=3290605.3300301.

[39] John C. McClelland, Robert J. Teather, and Audrey Girouard. 2017. Haptobend: shape-changing passive haptic feedback in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (SUI '17). Association for Computing Machinery, Brighton, United Kingdom, 82–90. https://doi.org/10.1145/3131277.3132179.

[40] Tim McGraw, Esteban Garcia, and Drew Sumner. 2017. Interactive swept surface modeling in virtual reality with motion-tracked controllers. In *Proceedings of the Symposium on Sketch-Based Interfaces and Modeling* number 1. ACM, New York, NY, USA, (July 2017), 1–9. https://dl.acm.org/doi/10.1145/3092907.3092908.

[41] Mark Mine, Arun Yoganandan, and Dane Coffey. 2014. Making VR Work: Building a Real-World Immersive Modeling Application in the Virtual World. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*. ACM, New York, NY, USA, (October 2014), 80–89. https://dl.acm.org/doi/10.1145/2659766.2659780.

[42] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, C. Schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions*, 21, 3, (May 2014), 40–45. http://dl.acm.org/citation.cfm?doid=2608008.2591689.

[43] Thomas Muender, Anke V Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. Does It Feel Real? Using Tangibles with Different Fidelities to Build and Explore Scenes in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–12. http://dl.acm.org/citation.cfm?doid=3290605.3300903.

[44] Miguel A Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, New York, New York, USA, 1099. http://dl.acm.org/citation.cfm?doid=2470654.2466142.

[45] Vijayakumar Nanjappan, Hai-Ning Liang, Feiyu Lu, Konstantinos Papangelis, Yong Yue, and Ka Lok Man. 2018. User-elicited dual-hand interactions for manipulating 3d objects in virtual reality environments. *Hum.-Centric Comput. Inf. Sci.*, 8, 1, Article 154, (December 2018), 16 pages. https://doi.org/10.1186/s13673-018-0154-5.

[46] Donald A. Norman. 2002. *The Design of Everyday Things*. Basic Books, New York, NY, USA, 257.

[47] Erica L. Olmsted-Hawala, Elizabeth D. Murphy, Sam Hawala, and Kathleen T. Ashenfelter. 2010. Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). Association for Computing Machinery, Atlanta, Georgia, USA, 2381–2390. https://doi.org/10.1145/1753326.1753685.

[48] Claudio Pacchierotti, Stephen Sinclair, Massimiliano Solazzi, Antonio Frisoli, Vincent Hayward, and Domenico Prattichizzo. 2017. Wearable Haptic Systems for the Fingertip and the Hand: Taxonomy, Review, and Perspectives. *IEEE Transactions on Haptics*, 10, 4, (October 2017), 580–600. https://ieeexplore.ieee.org/document/7922602/.

[49] J. Perret and E. Vander Poorten. 2018. Touching virtual reality: a review of haptic gloves. In *ACTUATOR 2018; 16th International Conference on New Actuators*. VDE, Bremen, Germany, 270–274. https://ieeexplore.ieee.org/document/8470813.

[50] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. Volume 2013-April. ACM Press, New York, New York, USA, 955. http://dl.acm.org/citation.cfm?doid=2468356.2468527.

[51] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. 2002. Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9, 3, (September 2002), 171–193. https://doi.org/10.1145/568513.568514.

[52] M. H. Quenouille. 1949. Problems in Plane Sampling. *The Annals of Mathematical Statistics*, 20, 3, (September 1949), 355–375. http://projecteuclid.org/euclid.aoms/1177729989.

[53] Patrick Reipschläger and Raimund Dachselt. 2019. DesignAR: Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ACM, New York, NY, USA, (November 2019), 29–41. https://dl.acm.org/doi/10.1145/3343055.3359718.

[54] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). Association for Computing Machinery, Vancouver, BC, Canada, 197–206. https://doi.org/10.1145/1978942.1978971.

[55] Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. Volume 2015-April. ACM Press, New York, New York, USA, 3347–3350. http://dl.acm.org/citation.cfm?doid=2702123.2702583.

[56] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The Role of Physical Props in VR Climbing Environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* number Chi. ACM Press, New York, New York, USA, 1–13. http://dl.acm.org/citation.cfm?doid=3290605.3300413.

[57] William A. Scott. 1955. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*, 19, 3, 321. https://academic.oup.com/poq/article-lookup/doi/10.1086/266577.

[58] Jotaro Shigeyama, Takeru Hashimoto, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2019. Transcalibur: AWeight Shifting Virtual Reality Controller for 2D Shape Rendering based on Computational Perception Model. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–11. http://dl.acm.org/citation.cfm?doid=3290605.3300241.

[59] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. Volume 2015-April. ACM Press, New York, New York, USA, 3307–3316. http://dl.acm.org/citation.cfm?doid=2702123.2702389.

[60] Mike Sinclair, Eyal Ofek, Mar Gonzalez-Franco, and Christian Holz. 2019. CapstanCrunch: A Haptic VR Controller with User-supplied Force Feedback. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, (October 2019), 815–829. https://dl.acm.org/doi/10.1145/3332165.3347891.

[61] Alexa F. Siu, Eric J. Gonzalez, Shenli Yuan, Jason B. Ginsberg, and Sean Follmer. 2018. shapeShift: 2D Spatial Manipulation and Self-Actuation of Tabletop Shape Displays for Tangible and Haptic Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Volume 2018-April. ACM Press, New York, New York, USA, 1–13. http://dl.acm.org/citation.cfm?doid=3173574.3173865.

[62] Evan Strasnick, Christian Holz, Eyal Ofek, Mike Sinclair, and Hrvoje Benko. 2018. Haptic links: bimanual haptics for virtual reality using variable stiffness actuation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3174218.

[63] Keisuke Suzuki, Sohei Wakisaka, and Naotaka Fujii. 2012. Substitutional Reality System: A Novel Experimental Platform for Experiencing Alternative Reality. *Scientific Reports*, 2, 1, (December 2012), 459. http://www.nature.com/articles/srep00459.

[64] Shan-yuan Teng, Cheng-Lung Lin, Chi-huan Chiang, Tzu-Sheng Kuo, Liwei Chan, Da-Yuan Huang, and Bing-Yu Chen. 2019. TilePoP: Tile-type Pop-up Prop for Virtual Reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, (October 2019), 639–649. https://dl.acm.org/doi/10.1145/3332165.3347958.

[65] Theophanis Tsandilas. 2018. Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation. *ACM Transactions on Computer-Human Interaction*, 25, 3, (June 2018), 1–49. https://dl.acm.org/doi/10.1145/3182168.

[66] Consuelo Valdes, Diana Eastman, Casey Grote, Shantanu Thatte, Orit Shaer, Ali Mazalek, Brygg Ullmer, and Miriam K. Konkel. 2014. Exploring the design space of gestural interaction with active tokens through user-defined

gestures. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, New York, New York, USA, 4107–4116. http://dl.acm.org/citation.cfm?doid=2556288.2557373.

[67] Radu-Daniel Vatavu. 2012. User-Defined Gestures for Free-Hand TV Control. In *Proceedings of the 10th European conference on Interactive tv and video - EuroiTV '12*. ACM Press, New York, New York, USA, 45. http://dl.acm.org/citation.cfm?doid=2325616.2325626.

[68] 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, Seoul, Republic of Korea, 1325–1334. https://doi.org/10.1145/2702123.2702223.

[69] Christian Weichel, Manfred Lau, David Kim, Nicolas Villar, and Hans W. Gellersen. 2014. Mixfab: a mixed-reality environment for personal fabrication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). Association for Computing Machinery, Toronto, Ontario, Canada, 3855–3864. https://doi.org/10.1145/2556288.2557090.

[70] Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. Haptic Revolver: Touch, Shear, Texture, and Shape Rendering on a Reconfgurable Virtual Reality Controller. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Volume 2018-April. ACM, New York, NY, USA, (April 2018), 1–4. https://dl.acm.org/doi/10.1145/3170427.3186515.

[71] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. Volume 8. ACM Press, New York, New York, USA, 1083. http://dl.acm.org/citation.cfm?doid=1518701.1518866.

[72] 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 '05*. ACM Press, New York, New York, USA, 1869. http://portal.acm.org/citation.cfm?doid=1056808.1057043.

[73] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. Virtualgrasp: leveraging experience of interacting with physical objects to facilitate digital object retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3173652.

[74] Jackie (Junrui) Yang, Hiroshi Horii, Alexander Thayer, and Rafael Ballagas. 2018. Vr grabbers: ungrounded haptic retargeting for precision grabbing tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (UIST '18). Association for Computing Machinery, Berlin, Germany, 889–899. https://doi.org/10.1145/3242587.3242643.

[75] Andre Zenner and Antonio Kruger. 2017. Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 23, 4, (April 2017), 1285–1294. https://ieeexplore.ieee.org/document/7833030/.

[76] André Zenner and Antonio Krüger. 2019. Drag:on– A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–12. http://dl.acm.org/citation.cfm?doid=3290605.3300441.

[77] Kening Zhu, Taizhou Chen, Feng Han, and Yi-Shiun Wu. 2019. HapTwist: Creating Interactive Haptic Proxies in Virtual Reality Using Low-cost Twistable Artefacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–13. http://dl.acm.org/citation.cfm?doid=3290605.3300923.

Received July 2021; revised September 2021; accepted September 2021