Hands-on, Hands-off: Gaze-Assisted Bimanual 3D Interaction

Mathias N. Lystbæk
Aarhus University, Denmark
mathiasl@cs.au.dk

Thorbjørn Mikkelsen
Aarhus University, Denmark
201905196@post.au.dk

Roland Krisztandl
Aarhus University, Denmark
202202868@post.au.dk

Eric J. Gonzalez
Google, U.S.A.
ejgonz@google.com

Mar Gonzalez-Franco
Google, U.S.A.
margon@google.com

Hans Gellersen
Lancaster University, U.K.
Aarhus University, Denmark
h.gellersen@lancaster.ac.uk

Ken Pfeuffer
Aarhus University, Denmark
ken@cs.au.dk

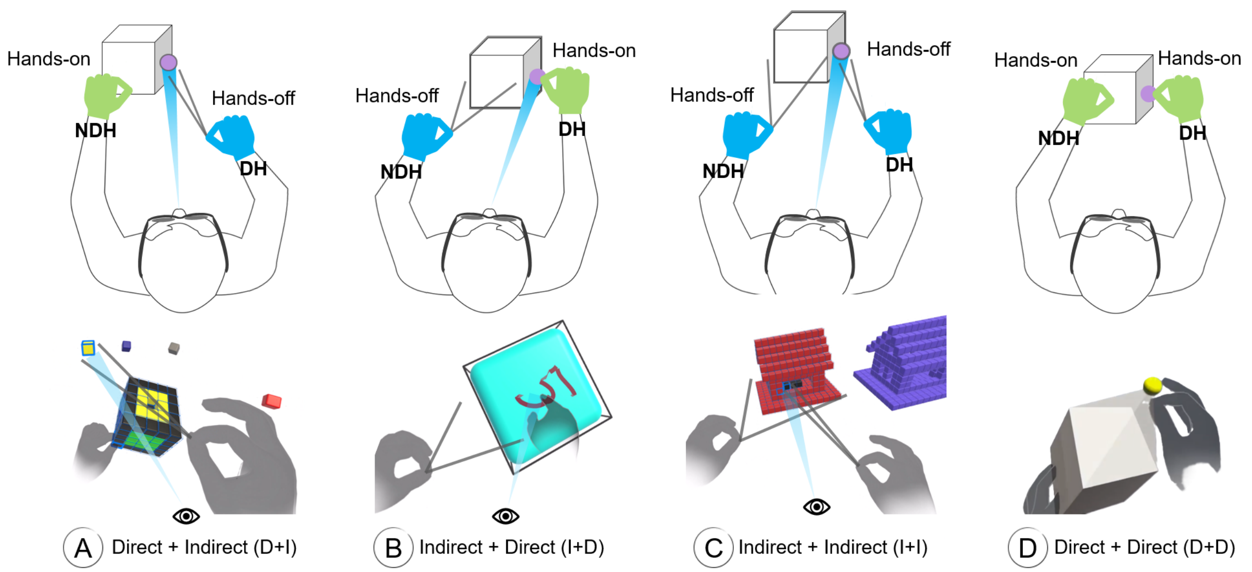


Figure 1: Eyes and hands in 3D enable novel, fluid interplays between our two hands. With each dominant (DH) and non-dominant hand (NDH), we can transition between direct or indirect gestures via Gaze+Pinch. E.g., to hold a model and rapidly pull and attach parts to it (A). Conversely, to indirectly manipulate it without occlusion for clear DH input (B). Dual-indirect modes offer occlusion-free interaction with low effort (C). All are complementary options to standard direct manipulation (D).

ABSTRACT

Extended Reality (XR) systems with hand-tracking support direct manipulation of objects with both hands. A common interaction in this context is for the non-dominant hand (NDH) to orient an object for input by the dominant hand (DH). We explore bimanual interaction with gaze through three new modes of interaction where the input of the NDH, DH, or both hands is indirect based on Gaze+Pinch. These modes enable a new dynamic interplay between our hands, allowing flexible alternation between and pairing of complementary operations. Through applications, we demonstrate several use cases in the context of 3D modelling, where users exploit occlusion-free, low-effort, and fluid two-handed manipulation. To gain a deeper understanding of each mode, we present a user study on an asymmetric rotate-translate task. Most participants preferred indirect input with both hands for lower physical effort, without a penalty on user performance. Otherwise, they preferred modes where the NDH oriented the object directly, supporting preshaping of the hand, which is more challenging with indirect gestures. The insights gained are of relevance for the design of XR interfaces that aim to leverage eye and hand input in tandem.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10.
https://doi.org/10.1145/3654777.3676331

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Virtual reality**; **Gestural input**.

KEYWORDS

virtual reality, eye-tracking, gaze input, bimanual interaction, 3D manipulation

ACM Reference Format:
Mathias N. Lystbæk, Thorbjørn Mikkelsen, Roland Krisztandl, Eric J. Gonzalez, Mar Gonzalez-Franco, Hans Gellersen, and Ken Pfeuffer. 2024. Hands-on, Hands-off: Gaze-Assisted Bimanual 3D Interaction. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3654777.3676331

1 INTRODUCTION

Extended Reality (XR) systems with hand-tracking support direct manipulation of 3D objects in near space without the need for controllers. As in physical reality, object manipulation in XR frequently involves both hands in asymmetrical roles [48]. A common pattern is for one hand to assist the other by orienting an object for manipulation [13, 15]. For example, we might turn a cup with one hand to make it easier for the other to grasp the handle or rotate a model in one hand to attach parts in different places to it. The foundations of bimanual-asymmetric interaction have been widely studied for hand-controlled inputs, with the non-dominant hand (NDH) setting the spatial frame of reference for actions of the dominant hand (DH) [9]. In this work, we explore how bimanual interactions of this type can be assisted by gaze.

Gaze lends itself naturally to extending manual input. Users can acquire objects more quickly by gaze than with their hands, and gestures performed with the hand can be applied on the gaze target [28]. Commonly, this is motivated to extend reach to far space in XR, for manipulation of objects that are beyond the direct input range of the hands [5, 32, 41]. However, work on touch interfaces has shown advantages of using gaze also closer up, for example, to reduce effort in repetitive input and to manipulate objects without occluding them [28, 30, 31]. For this work, we are inspired by techniques such as *Gaze-Shifting*, where gaze supports manual input dynamically in different modes: (i) for direct input when both eyes and hands are on the same object, and (ii) for indirect input when the hands are off-target with respect to gaze [29]. Applied to bimanual-asymmetric interaction in 3D, we can then consider "hands-on" versus "hands-off" input by either hand.

Figure 1 illustrates gaze-assisted bimanual-asymmetric interaction with three new modes of interaction, demonstrated in 3D modelling applications. In D+I mode (Figure 1A), input by the NDH is direct, and input by DH is indirect by Gaze+Pinch [32]. This mode is useful, for example, in assembly operations, where the NDH holds the model to orient it for attachment of parts, while the DH is assisted by gaze to reach for parts to pick and attach, without having to move much and without having to cross the other hand. In I+D mode (Figure 1B), the roles are reversed. This is useful, for instance, when the DH is used for drawing or writing on the surface of a 3D object, with the NDH used indirectly to orient the object for the task without occluding it. In I+I mode (Figure 1C), both hands operate indirectly. A use case for this is tasks that involve switching between objects that are manipulated, for example, to extract parts from one object to insert into the other, we also see accessibility applications being enabled by the dissociation of direct interaction of our techniques.

To gain a deeper understanding of the strength and limitations, we conducted a user study ($N = 16$) in which we compared D+I, I+D, and I+I with direct manipulation by both hands on the object (D+D) as baseline. The bimanual task required orienting a cube by the NDH for a dragging interaction by the DH on different faces of the cube. The conditions used pinch as the selection mechanism and differed in whether the pinch was performed directly "hands-on", or indirectly "hands-off" based on Gaze+Pinch.

The results show that indirection of input significantly lowers physical effort in near space, without a penalty on overall task completion time. However, we note that users were faster in the subtask of object acquisition with gaze, but slower in the subsequent setting of the frame of reference for the DH's action. Meaning, users had more difficulty rotating an object indirectly, as this lacks the natural ways of preshaping the hand and wrist to optimize for the required rotation after object acquisition. This issue was mainly noticeable with I+D, where the NDH's rotation difficulty affected the spatial coordination with the DH's direct input. In contrast, fully indirect control (I+I) avoided this problem by eliminating direct contact through gaze-based selection. This, plus the results of the least physical effort, led to the most preferred technique (9/16 users) for this task. The second preference was D+I (4/16), which better supports natural preshaping and achieved similarly good overall performance. However, I+I remained preferred overall due to its consistency across hands.

In sum, this is a first work to consider gaze specifically in support of bimanual interactions in reachable space. The main contribution lies in the exploration through applications and comparative evaluation of three modes enabled by gaze, with either or both hands used indirectly for orienting and manipulation of objects in 3D. The work demonstrates benefits of using Gaze+Pinch over direct manipulation and gives insight into combinations of direct versus indirect input, including challenges of preshaping when hand input is indirect and mediated by gaze. While the work focuses on comparison of direct and gaze-assisted modes, they are all supported within a framework of "hands-on, hands-off" interaction in which users can dynamically transition from one mode to another.

2 RELATED WORK

Our work intersects the research spaces of bimanual and gaze in human-computer interaction (HCI).

2.1 Bimanual Interaction

From writing to 3D modelling, many tasks involve coordinated bimanual interaction in an asymmetrical way. Guiard describes three key principles in human skilled bimanual action: (1) The NDH sets the spatial frame for actions of the DH, (2) the NDH precedes the DH in action, and (3) each hand operates in different scales of motion, with the NDH performing coarse irregular actions and the DH performing detailed and frequent ones [9]. In the HCI field, these principles inform interaction design to effectively devise two-handed user interfaces [2, 3, 9, 13–15, 21].

In virtual environments, there is a rich body of research on one-handed and bimanual-symmetric constellations, focusing on rotation and manipulation tasks individually [8, 40, 46] or in parallel in docking tasks [20, 43, 49]. Hinckley et al. conducted a cooperative bimanual interaction study where the DH manipulates a tool while the NDH holds the target object [15]. The study showed that performance is best in more complex tasks with this division of labour across hands, aligning with Guiard's principles. In our study, we design a similar abstracted task in a virtual environment.

Researchers have widely explored 3D UIs that use tracked controllers and hand gestures for manipulating objects in virtual environments [2, 6, 12, 44]. For instance, early work by Cutler et al. explored a range of immersive 3D modelling interactions with 6DOF controllers, such as grab-and-scale, free/axis/heuristic rotation, constrained translation, and many more, demonstrating the wide utility of asymmetric interaction options [6]. Recent examples, such as TabletInVR [38] and VRSketchIn [7], explore bimanual applications integrating a tablet into 3D. This setup facilitates asymmetrical interaction, with the NDH holding the tablet as a spatial reference for the DH, useful for precise sketching with touch inputs or a stylus, miniature world control, object cutting, and gestural commands centred around the tablet. These applications can in principle include eye-tracking-based input enhancements, for which it is of interest to study how the eyes and two hands compare to the direct manipulation of two-handed gestures. Our study task is similarly designed with the NDH setting the spatial reference frame and the DH manipulating parts that are within that reference frame.

2.2 Gaze and Hand Interaction

Using the eyes to point and select virtual elements in the user interface can be a fast, natural, and convenient input method [17]. However, a key challenge is Midas touch, i.e., whether a gaze at an object is meant as an intended button selection or a simple glance of the user [16]. Dwell-time can be used to alleviate this issue, and a dedicated manual trigger based on our hand allows conflict-free looking and selection. In our work, we use a pinch gesture to confirm selection, based on the Gaze+Pinch technique [32].

Several works have explored how the use of manual triggers can extend to various manipulation tasks enabled by gaze and touch inputs [28, 37, 39]. A common theme is *gaze selects, hand manipulates*, which allows redirecting hand manipulations to the object the user is looking at, rendering the gestures indirect [28]. This concept was pushed further through the work on *Gaze-Shifting* [29], the principle idea being that interfaces can support both direct and indirect gestures, with a seamless transition mechanism based on the user's natural eye-hand coordination. This allows for new interaction benefits, as users can capitalise on the contrasting properties of direct/indirect by switching between them or using them simultaneously via two hands in pen-and-touch systems (using the pen in direct and touch in indirect modes) [30] and two-handed tablet interaction (the NDH holds the device, the DH enters inputs [31]). In our work, we take inspiration from *Gaze-Shifting* to enable seamless transitioning between direct and indirect interaction based on hand proximity to gazed-upon objects.

In XR, gaze pointing is established, for eyes-only and multimodal interaction through coupling with controller [19, 33, 50] and hand-tracking [22, 23, 25, 32, 42]. For two-handed interaction, one study evaluated the bimanual-asymmetric task of 3D sketching with the DH and mode-switching in a colour palette menu held in the NDH [33]. Several techniques were investigated that employed gaze and controller inputs in the DH to select the colour in the palette. Direct manipulation has been found as the best-performing technique, but a general trend showed that eye-hand techniques lowered the physical effort, although they traded it with performance, and fully eye-based techniques led to eye fatigue. However, the study focused on using controllers for a menu task, a specialised design task. In our work, we explore the fundamental effects of integrating eye-gaze in a cooperative bimanual-asymmetric task.

3 HANDS-ON, HANDS-OFF INTERACTION

For example, the hand holding the object can rotate the object such that a desired part of the object is in view, thereafter the other hand can interact with and manipulate the desired part. We explore the role and needs for spatial interaction of each of the hands through the "*Hands-on, Hands-off*" framework: It encapsulates the bimanual input modes and their transitions. Specifically, three combinations of direct "hands-on", and indirect "hands-off" bimanual interaction that is gaze-assisted (Figure 1):

D+I Directly framing + Indirectly manipulating the object (A)
I+D Indirectly framing + Directly manipulating the object (B)
I+I Indirectly framing + Indirectly manipulating the object (C)

In the following, we detail transitions between modes, interaction properties, and implementational details.

3.1 Transitioning

Both D+I and I+D allow users to leverage the advantages of both direct and indirect interaction however they desire in a given moment; they should be able to freely transition from one to the other, or even to I+I or D+D. To facilitate this seamless interaction, it is essential that users can interact how they prefer in the moment. If switching from direct to indirect, or vice versa, is cumbersome then users may become frustrated and experience difficulty in their current task. We take inspiration from [29] to transition based on the distance from the user's hand to the gazed-upon object; allowing users to seamlessly transition between direct and indirect interactions, based on whether the object is in hand reach.

This transition also allows for flexibility in which hand does what. If the user desires to hold an object directly in their right hand, they simply need to reach out to grab it; if indirect is desired, the user simply needs to keep their hand slightly further away, likely in a more comfortable position.

3.2 Interaction Properties

We analyse relevant interaction properties from the unity of direct and indirect inputs for both hands. Table 1 provides an overview of some of the characteristics of each interaction combination.

3.2.1 *Input Classes.*
There are two main classes of input for technology: direct manipulation that mimics real-world actions, and indirect interaction that uses offset hand positions [11, 34].

	D+D	D+I	I+D	I+I
Indirection	None	Low-Medium	Low-Medium	High
Occlusion	High	Medium	Medium	Low
Simul. selection	Yes	Yes	Yes	No

Table 1: Interaction suitability overview.

Direct manipulation allows users to interact with objects in the way they are accustomed to in the real world. Users can utilise their proprioception along with visual cues of how their hand is moving relative to the object they are manipulating.

Indirect manipulation allows users to interact over distance, with less physical effort by using relative control-display mappings. Using gaze, users can simply look at the object they wish to interact with, then "grasp" it through an indirect pinch gesture [32], after which the user's hand movement and rotation is mapped to the interacted object.

3.2.2 Occlusion. Occlusion caused by the user's hands can be problematic in mainly two ways: (1) the user's hands might occlude some parts which the user wishes to interact with, slowing down the user, and (2) one of the user's hands might occlude the other, which is problematic with vision-based hand-tracking systems[1] as it will likely cause diminished hand-tracking quality or even complete hand-tracking loss. Indirect interaction mitigates this issue as the user's hands are naturally able to be further apart and the user does not need to cross their arms to reach parts opposite the other (e.g., holding something with the left hand while trying to grab something on the left with the right hand).

However, this occlusion can also be leveraged for good, as a visual aid. When the user's hand is directly on the object, the user is afforded more visual feedback on how their hand is rotating and moving in direct relation to how the object is moving. For example, the user may more easily be able to rotate an object on a particular axis because they can see how their hand is rotating and how it is influencing the object, then, if the rotation seems off, quickly correct the rotation through intuition.

3.2.3 Simultaneous Selection. Both D+I and I+D offer the same ability as D+D; users are able to acquire two objects at the same time. The user can acquire an object indirectly by gazing at the object and pinching the indirect hand, while, at the same time, acquiring something with their direct hand that they do not need to look at. I+I lacks this ability as it is limited by only one gaze pointer. As such, if users wish to acquire multiple objects, they need to first look at one, perform a pinch with one hand, then look at another and perform another pinch with the other hand. This limits the acquisition of multiple objects with I+I to be sequential. However, once objects have been acquired, they can be freely rotated and manipulated in parallel, without users needing to worry about where they are looking.

3.3 Implementation Details

As hand-tracking has some degree of jitter and noise, we filter the hand-tracking signal using the 1€ filter [4], smoothing out smaller, high-frequency movements ($mincutoff = 0.5$) while keeping larger, quick movements, allowing for snappy motion ($\beta = 5$).

[1]E.g., the Microsoft HoloLens 2, Meta Quest Pro, and Apple Vision Pro.

Eye-tracking can be similarly noisy and could therefore benefit from filters. However, these filters can cause cases where the gaze signal is significantly lagging behind the user's true gaze, if not tuned correctly. As such, in our work, we keep the raw, unfiltered gaze signal. To compensate, we set the target scale to be sufficient for gaze selection to avoid errors for the system we used.

To allow users to understand which objects can be indirectly manipulated, we highlight such objects when the user's gaze falls upon them. The highlighting is presented as a grey additional outline around the object, meaning that the object itself appears unchanged. To visually communicate indirect control, we add two grey lines from the user's thumb to the edges of the object, while the outline remains. The object appears to be within a funnel, as if captured by the user (cf. Figure 1A-C). When the user's hand is within 5 cm of an object, the hand's input transitions to direct interaction, and the gaze highlighting is disabled to not distract the user as their hand should serve as enough visual feedback (cf. Figure 1A+B+D).

We provide our implementation as an open-source tool[2] released as a Unity package built around the Meta XR-all-in-one SDK v62[3]. The Unity package allows developers to simply import the package, after having imported and set up the Meta SDK, add a prefab to their scene and add one component to objects they wish to interact with (requiring the object to have a Meta "`HandGrabInteractable`" component for direct interaction).

4 APPLICATIONS

We explore examples for the investigated interactions through an application probe in the domain of 3D modelling with direct manipulation, inspired by prior papers [6, 7, 12, 32, 38]. The application supports the framework with all four modes and transitions between them. The application provides a menu, which users can access by holding up their NDH. The menu contains an object library, various tools, and settings. One feature is the colouring tool, where users can colour objects in the virtual environment. They can choose from a variety of colours from the menu. A second feature is snapping. If active, when two objects are close to each other, they will snap together automatically, or when the manipulation ends on one of them. Snapped objects can then be manipulated together. To remove an object from a snapped group, the user has to start manipulating one of the objects in the group, and then perform a (second) selection on the object that should be removed.

4.1 Shape Voxel Colouring: Selections with (D+I)

This application showcases benefits of directly grabbing a 3D object using the NDH to fluidly set the reference frame for the DH to perform successive gaze-based selections. The indirect manipulation on the DH enables quick and accurate selection of the voxels of the cube using gaze. In this application, the desired outcome is to colour them in a specific pattern. Figure 2 shows three steps of the process, from the initial stage of the cube to the user hovering over a voxel to be coloured, and finally applying the colour. The DH accurately rotates the cube before the DH colours desired parts of the cube.

[2]https://github.com/Matho97/hands-on-hands-off
[3]https://assetstore.unity.com/packages/tools/integration/meta-xr-all-in-one-sdk-269657 - last accessed April 2nd, 2024.

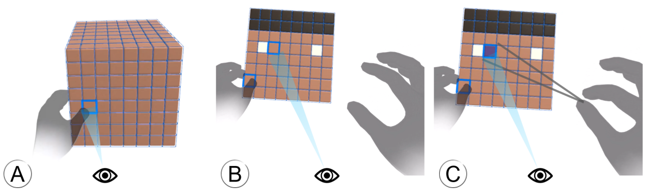


Figure 2: D+I Voxel Colour application: (A) Acquire the big voxel cube with NDH (direct). (B) Look at a voxel and pinch with DH to colour (indirect). (C) Colour multiple voxels quickly and accurately via gaze.

4.2 Mesh Manipulation: Coordinated Dragging (D+I)

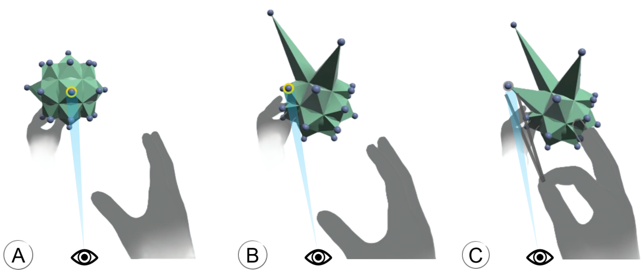


Figure 3: D+I Moravian Star application: (A) Acquire the Moravian Star with NDH (direct). (B) Look at a star point and pinch with DH to extrude (indirect). (C) Avoids hand-tracking camera occlusion for smoother operation.

As in the prior example, this application showcases the benefit of intuitively setting the model's frame of reference directly with the NDH, to aid the DH's task. Different to the prior task, here we demonstrate how users can rapidly fire off Gaze+Pinch based dragging operations on the model, coordinated with the NDH holding the object. We demonstrate this capability on the example of modelling a Moravian Star shape, which has 26 points in a generic shape. Figure 3 shows three steps from the manipulation of the Moravian Star, from starting with shorter points to manipulating the points to form a Moravian Star mesh. With continuous manipulations, each part of the star adds more occlusion to the task, making it more difficult to reach and manipulate the next part. Yet, the indirect interaction with the DH mitigates issues of occlusion. If both hands would directly operate on the model, occlusion is amplified, and, from a technical perspective, there is potential loss of tracking where one hand would occlude the other resulting in lost tracking. This is resolved by using one hand in indirect mode.

4.3 Disassembly: Independent Dragging (D+I)

As the third D+I example on direct NDH orientation, this example is distinct in the DH's action that is independent of the NDH after target acquisition. The user brings into view parts that need to be removed and selects the object with Gaze+Pinch of the DH, which detaches the object from the model. Then, the user can continue to move the object independently, while the NDH can separately manipulate the model in hand. Figure 4 shows three steps of the process, from the initial stage of the pyramid with highlighted parts, to the user hovering over a cube to remove it. Thus, unlike previous examples, we demonstrate that users can swiftly work with independent objects, extract parts from a model, or move distinct objects inside as needed.

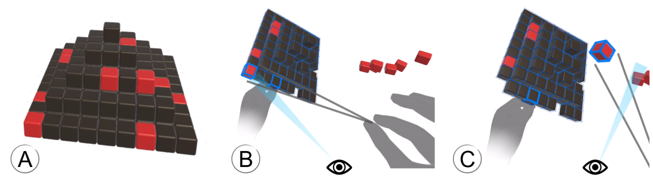


Figure 4: D+I Dissasembly application: (A) Acquire the pyramid of cubes with the NDH (direct). (B) Look at a red cube and pinch with DH to remove it (indirect). (C) Fast, accurate removal of cubes without issues of occlusion.

4.4 Drawing on a Cube (I+D)

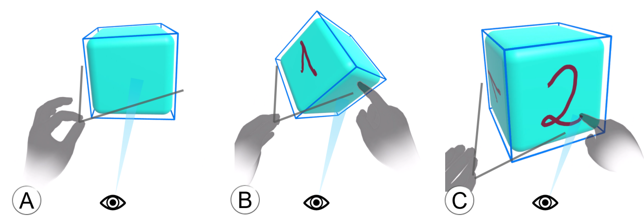


Figure 5: I+D Cube Drawing application: (A) Gaze at the cube and pinch with NDH (indirect). (B) Rotate the cube (indirect) and (C) draw on the desired side with DH (direct).

This application, based on drawing numbers on the sides of a cube, showcases the benefit of occlusion-free reference setting, afforded by the indirect NDH, combined with the benefit of precise surface interaction, afforded by the direct DH. Figure 5 shows three steps of the process, from the initial stage of the cube to the user rotating the cube to the desired side to draw on, and finally drawing the number. The user can rotate the cube without worrying about occluding a side with their hand while drawing, mitigating the need to clutch the cube.

4.5 Extraction of Small Objects (I+I)

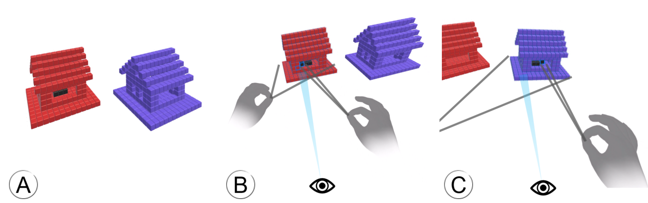


Figure 6: I+I Small Object Extraction application: (A) Two houses with small objects within. (B) After gazing at the red house and pinching with NDH (indirect), then rotating the red house to see inside it, gaze at a small object inside it and pinch with DH (indirect). (C) Quickly select the other, purple house with NDH (indirect) and place the small object through the window.

Figure 7: Trial sequence with the D+D interaction on the *Single* task: (A) Cube starts in default position. (B) NDH acquires (blue point appears). (C) Cube is prepared for DH manipulation. (D) DH manipulates point to target. (E) DH releases, ending trial.

This application, based on removing some objects from an already existing larger build and adding them into another build, showcases how fully indirect interaction can still support simultaneous manipulation, through a sequence of Gaze+Pinch selections. Figure 6 shows three steps of the process, from the initial stage of the two large builds to the user removing an object from the red build, and finally adding it to the purple build. This application benefits from indirectly grabbing the larger building, allowing the user to rotate it such that the user can easily see the object that has to be removed or the location where the new objects should be placed. It also leverages indirect manipulation of the small objects, as direct manipulation would be cumbersome due to the small size of the objects and would also result in the hand occluding parts of the build and accidental selections of other objects.

5 USER STUDY

We evaluate D+I, I+D, I+I, and D+D in a user study of a bimanual-asymmetric task to understand user performance and experience. The task involves the user re-orienting an object with their NDH, and then using their DH to manipulate a part connected to the object. Our main research questions are:

- RQ1: **Eye-hand vs. Direct Hand:** How does gaze-assisted input compare to direct manipulation? Indirect input might reduce effort but direct manipulation can be intuitive.
- RQ2: **D+I vs. I+D:** How does the D+I and I+D mapping for DH and NDH affect the task performance? D+I can benefit from the NDH setting the reference frame with rotation, but I+D could be advantageous for familiar DH manipulation.
- RQ3: **D+I and I+D vs. I+I:** How does two-handed indirect input (I+I) compare to partially indirect modes (D+I and I+D) in terms of user performance (speed, error-rate), perceived effort, and usability? I+I might have lower effort and higher usability, but hybrid combinations could offer a new balance between intuitiveness, speed, and ease of use.
- RQ4: **Manipulation patterns for indirect input:** How do users accomplish rotation and dragging tasks with eye-hand input, considering the lack of physical affordances that are present in direct manipulation?

5.1 Task Design

The task involves the acquisition of a cube and subsequent manipulation of a part of that cube. Figure 7 illustrates an example trial completion; first, the user (1) acquires the cube with the NDH, (2) prepares the cube with the NDH by rotating it, for later manipulation, (3) manipulates a sphere on the cube with the DH, and (4) releases the DH pinch to finish.

The cube is 12.5 cm and appears 50 cm in front of the participant's eye level. The "target" is a 2.5 cm-diameter green sphere located in the centre of a cube's face, appearing 12.5cm from the side(s) of the cube, positioned relative to the cube. After grabbing the cube, a new blue "part" sphere of the same size is placed at the surface of the cube, which the user needs to extrude to the green "target" sphere to complete the task. Participants can clutch the NDH's rotation while performing the task, but manipulation with the DH can only happen once. The manipulation is 1D, meaning the part can only move in a straight line towards the green target sphere. This is because we are interested in the two-handed coordination, while the manipulation is secondary.

We test two task variations. The ***Single*** **task** focuses on a rapid, one-off acquire-prepare-manipulate sequence (cf. Figure 7). At each trial, only one of the six sides of the cube will appear with a green target sphere. The participant would then conduct the task. Once the participant finishes a trial, the system waits for one second before the cube resets to 50 centimetres in front of the participant at eye level while also resetting the blue and green spheres. This represents a fundamental task with high internal validity. The ***Continuous*** **task** focuses on a continuous acquire-prepare-manipulate sequence, introducing carryover and clutching behaviour (cf. Figure 8). All six sides of the cube have visible green target spheres. Once the participant interacts with the cube using their NDH, all the targets appear and the task finishes after all six faces have been extruded. We hide the completed blue and green spheres when the participant completes a trial to indicate that they have completed that side.

The *Continuous* task is important and distinct from the *Single* task as it captures a more realistic interaction behaviour. When manipulating a single part, the previously manipulated part may be occluding the current part. Further, users may plan ahead to manipulate multiple parts, for a more efficient interaction over multiple steps. This can, e.g., affect how users coordinate their eye and hand movements around the working model.

5.2 Study Design

The within-subject study has one independent variable *Technique* with four conditions (order is counterbalanced). The *Single* task always precedes the *Continuous* task to present increasing difficulty.

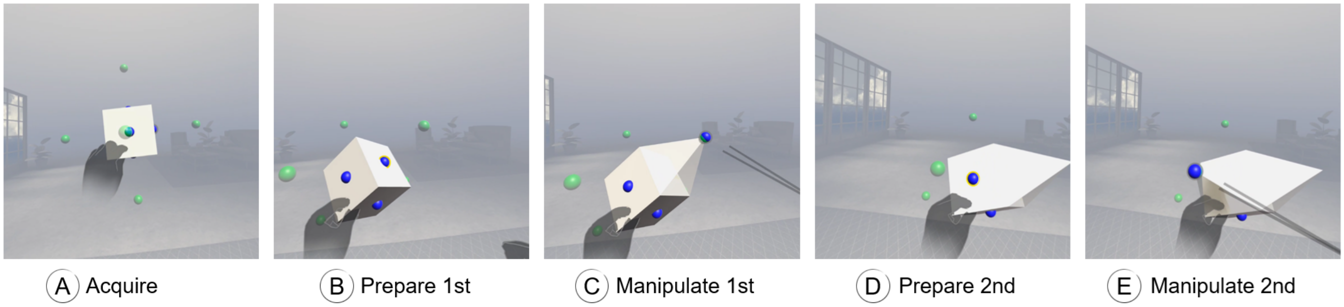


Figure 8: Example trials in the *Continuous* task showcased with D+I interaction. The *Continuous* task starts as the *Single* task (A-C). However, in the *Continuous* task, the cube does not reset after each sphere manipulation, as in the *Single* task. Instead, the faces of the cube remain extruded (D-E) and the cube only resets after the final trial in the block (i.e. every six trials). As such, the user does not need to perform as many acquisitions, although the user may need to clutch (let go and re-acquire) to reach the other side of the cube.

In the *Single* task, we randomised the order of the target sides on the cube in each block of six trials. Participants completed four blocks of six trials, resulting in 4 *Techniques* ×2 *Tasks* ×4 *Blocks* × 6 *Cube Sides* = 192 trials per participant.

5.3 Apparatus and Implementation

The study is implemented with Unity (2022.3.25f1) for the Meta Quest Pro (90 Hz, 30 Hz eye tracker) using the Meta XR All-in-One SDK (v62). Eye-tracking accuracy varies from 1.5° to 3° [1, 45]. To mitigate hardware issues caused by eye-tracking inaccuracies, we increased the invisible selection radius of the cube (50% bigger than visual) and the sphere (300% bigger). As we are mainly interested in the bimanual coordination, rather than the selection of small targets, this allows us to ensure that typical issues with eye-tracking inaccuracy effects will be minimised. Furthermore, hand-jitter was filtered by the 1€ Filter [4].

5.4 Procedure

Participants were briefed and completed consent and demographics forms. Afterwards, two short videos of the two tasks for the starting technique were shown to understand the study elements. Participants then wore the headset and underwent fit adjustment and eye-tracking calibration. In the *Single* task, participants had two training blocks (12 trials)[4] and were instructed to be as fast and as accurate as possible. Subsequently, after completing the four study blocks for the *Single* task, participants completed four more study blocks for the *Continuous* task. After each condition, participants completed a post-condition questionnaire. After all conditions, the participants completed a final post-study questionnaire. The study lasted on average around 50 minutes.

5.5 Evaluation Metrics

The following data was collected during the study:

- *Acquisition Time*: The time from the start of a trial until the NDH acquired the cube.
- *Preparation Time*: Time from NDH acquiring the cube to DH acquiring the part.
- *Manipulation Time*: Time to manipulate the part until pinch-release for trial completion.
- *Trial Completion Time*: Time from start (cube appears) to end (DH pinch release).
- *Error Rate*: How often the final position of the part was further than 5cm from the green target sphere.
- *NDH, DH, and Combined Hand Movement*: Hand movement in each trial in meters recorded for both the NDH, the DH, and both Combined.
- *NASA-TLX*: Task load was measured after each condition using the NASA Task-Load Index questionnaire [10].
- *User Feedback*: After all study conditions were completed, the participants ranked the four interaction combinations in terms of preference and provided feedback about their choices.
- *Observations*: During the study, the experimenter observed participant behaviour.

5.6 Participants

We recruited 16 participants (10 male-identifying and 6 female-identifying) from the local university, consisting mainly of Computer Science researchers and Master's students. Participants' age ranged from 22 to 37 ($M = 26.13, SD = 3.95$). On a scale between 1 (low) and 5 (high), participants rated themselves as having average experience with VR/AR ($M = 2.69, SD = 1.08$) and 3D Hand Interaction ($M = 2.38, SD = 1.36$), while rating themselves as having little eye-gaze experience ($M = 1.44, SD = 0.51$). While the interaction combinations are handedness-agnostic, meaning that they work with either hand, we limited our study to right-handed participants to significantly simplify the study design and data analysis.

5.7 Results

We analysed the results via RM-ANOVA with post hoc tests. For non-normal and questionnaire data, we used the Friedman test with post hoc Wilcoxon tests. Both Holm-Bonferroni corrected. Statistical significance in graphs is shown as * for $p < .05$, ** for $p < .01$, and *** for $p < .001$ and error bars show 95% confidence intervals. We collected 3072 trials, of which 106 (≈3.45%) were discarded as errors (cf. Section 5.5). We also excluded 121 outliers from the analysis

[4]Indicated to be sufficient in pilot studies.

Single Task

Technique	Acquisition Time	Preparation Time	Trial Completion Time	NDH Movement	DH Movement	Combined Movement	Ranking
D+D	0.86 s	**0.97 s**	**3.03 s**	0.91 m	0.64 m	1.54 m	2.88
D+I	0.92 s	*1.09 s*	3.23 s	0.94 m	**0.44 m**	*1.38 m*	*2.38*
I+D	*0.82 s*	1.32 s	3.43 s	*0.70 m*	0.87 m	1.57 m	3.25
I+I	**0.73 s**	1.18 s	*3.21 s*	**0.57 m**	*0.52 m*	**1.09 m**	**1.50**
	$F_{3,45} = 10.52$	$F_{3,45} = 7.52$		$F_{3,45} = 17.49$	$F_{3,45} = 23.13$	$F_{3,45} = 13.03$	$\chi^2 = 16.5$
	$p < 0.001$	$p < 0.001$		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	$\omega^2 = 0.15$	$\omega^2 = 0.13$		$\omega^2 = 0.25$	$\omega^2 = 0.45$	$\omega^2 = 0.23$	$W = 0.34$

Continuous Task

Technique	Acquisition Time	Preparation Time	Trial Completion Time	NDH Movement	DH Movement	Combined Movement	Ranking
D+D	1.01 s	*1.23 s*	*2.4 s*	0.36 m	0.48 m	0.85 m	2.88
D+I	1.03 s	**1.10 s**	**2.32 s**	**0.31 m**	**0.38 m**	**0.69 m**	*2.38*
I+D	**0.86 s**	1.46 s	2.63 s	0.34 m	0.55 m	0.89 m	3.25
I+I	*0.87 s*	1.26 s	2.52 s	*0.31 m*	*0.43 m*	*0.74 m*	**1.50**
		$F_{1.89,28.39} = 6.16$			$F_{3,45} = 21.22$	$F_{3,45} = 10.72$	$\chi^2 = 16.5$
		$p = 0.007$			$p < 0.001$	$p < 0.001$	$p < 0.001$
		$\omega^2 = 0.1$			$\omega^2 = 0.34$	$\omega^2 = 0.18$	$W = 0.34$

Table 2: Significance statistics for both tasks across the measures, including post study rankings. The top mean for each measure is highlighted in bold font.

where Trial Completion Time exceeded $Mean \pm 3 \cdot SD$ (≈4.08% after excluding error trials).

Investigation on learning rates in Trial Completion Time showed no significant increase or decrease across trial blocks, consequently, we will analyse across all blocks. Table 2 shows an overview of the mean results, and the RM-ANOVA and Friedman analyses test results on measures where statistical significance was indicated.

5.7.1 Acquisition Time. Our results on Acquisition Time (Figure 9a) indicate that acquisition through gaze-assisted indirect interaction is faster than direct acquisition, which is to be expected in our tasks as the eyes move faster than the hands. In the *Single* task, our participants were significantly faster at acquiring the cube using the fully indirect I+I combination versus the two directly acquiring combinations D+D ($p = 0.002$) and D+I ($p < 0.001$).

5.7.2 Preparation Time. In terms of Preparation Time (Figure 9b), our results suggest that preparing (rotating) the cube is faster through direct interaction, possibly due to a greater sense of intuition. We found that, in the *Single* task, our participants were significantly quicker at preparing the cube for subsequent manipulation when using the fully direct D+D combination versus the two indirectly preparing combinations I+D ($p < 0.001$) and I+I ($p = 0.006$), while D+I was faster than I+D ($p = 0.02$). Similarly, in the *Continuous* task, D+I was significantly faster than I+D ($p < 0.001$).

5.7.3 Manipulation Time. We did not find any differences in Manipulation Time indicating that participants had no difficulties after having prepared the reference frame, regardless of interaction.

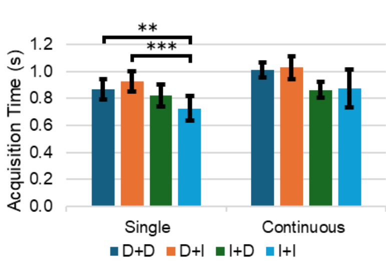
5.7.4 Trial Completion Time. Similarly, no significant differences were found in terms of Trial Completion Time (Figure 9c), indicating that the nuances of the interactions are more evident in acquisition (favouring gaze-assisted) and preparation (favouring direct) and mostly level out across the duration of the trial.

5.7.5 Error Rate. We found no significant differences in Error Rate in the *Single* task or the *Continuous* task.

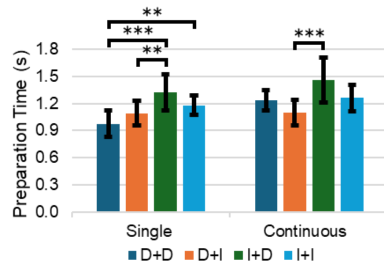
5.7.6 NDH Movement. Regarding NDH Movement (Figure 10a), our results indicate that direct interaction exerts more physical effort, as expected with our tasks requiring additional hand movement for directly acquiring the cube. In the *Single* task, our participants moved their hands significantly more with both D+D and D+I compared to both I+D and I+I (all $p < 0.004$).

5.7.7 DH Movement. In terms of DH Movement (Figure 10b), our participants exerted significantly more effort using I+D, even compared to D+D which both require direct manipulation using the DH, suggesting some fundamental difference in how the NDH set up the reference frame. In the *Single* task, our participants moved their DH significantly more with I+D compared to all other combinations (all $p < 0.001$), and significantly more with D+D compared to D+I ($p = 0.002$). For the *Continuous* task, our participants moved their DH significantly more with both D+D and I+D compared to both D+I and I+I (all $p < 0.043$), more with I+D compared to D+D ($p = 0.01$), and more with I+I compared to D+I ($p = 0.043$).

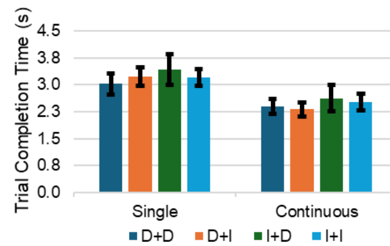
5.7.8 Combined Hand Movement. In terms of Combined Hand Movement (Figure 10c) it is clear that the indirect interactions exerted the least physical effort, as one would expect. However, the I+D combination did not exhibit this benefit, suggesting some difficulties compared to the other indirect combinations. We found that our participants moved their hands overall significantly less



(a) Acquisition Time.

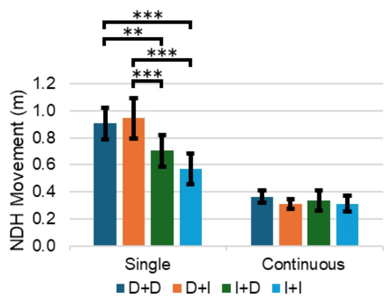


(b) Preparation Time.

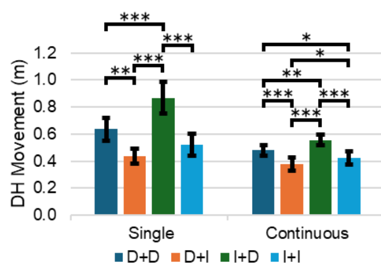


(c) Trial Completion Time.

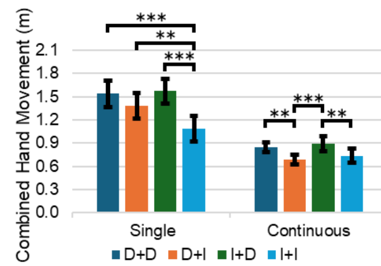
Figure 9: Results on (a) Acquisition Time, (b) Preparation Time, and (c) Trial Completion Time for each technique across tasks.



(a) NDH Movement.



(b) DH Movement.



(c) Combined Hand Movement

Figure 10: Results on (a) NDH Movement, (b) DH Movement, (c) Combined Hand Movement for each technique across tasks.

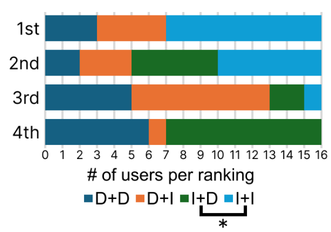


Figure 11: Results on user reported Rankings.

with I+I compared to all other interaction combinations (D+D, D+I, and I+D) (all $p < 0.006$). Similarly in the *Continuous* task; our participants moved their hands overall significantly more with I+D compared to both D+I ($p < 0.001$) and I+I ($p = 0.003$), participants also moved their hands overall more with D+D compared to D+I ($p = 0.002$).

5.7.9 User Feedback. We found no significant differences across any question on the NASA-TLX questionnaire.

Our results on reported rankings (Figure 11) indicate dissatisfaction with using indirect interaction for setting the reference, but mainly when the user also has to use direct interaction for manipulating within that reference frame. The frequencies for top rankings for techniques are: I+I (9 users), D+I (4), D+D (3). We found that I+I was significantly better ranked than I+D ($p < 0.001$).

The participants referred to the direct interactions as natural (3 participants) and easy (4 participants) but more fatiguing (9 participants), contrasting indirect interactions reported as less fatiguing (8 participants) and more comfortable (3 participants), but rotation was more difficult to comprehend (6 participants). Participants stated that having both hands in the same mode was easier to comprehend (5 participants), although participants also mentioned that both being direct sometimes caused crossing and tracking issues (2 participants).

5.7.10 Observations. The strategy employed for setting the reference frame varied greatly between participants. The mental rotation planning and execution also differed between techniques, while during-technique adjustments were mostly finished during training. We observed that participants struggled with the indirect rotation. Nine participants verbally mentioned that the cube rotated differently than planned for a given indirect input.

With direct input, participants would instinctively preshape their wrist, e.g., extending or flexing, allowing for greater rotation after acquisition, and limiting the need for clutching. However, the indirect approach affects this pattern, as all participants were observed acquiring the cube with a neutral wrist for I+D and I+I, effectively halving the rotational input domain. Interestingly, five participants became aware of this issue and attempted to improve, with three participants "re-learning" indirect preshaping in a short time.

Different usage strategies can be observed for every combination of technique, task, and target. These are observed as motor patterns informed by the current mental model and intention of the participant. It was more challenging when the manipulation was on the cube's left side. NDH preshaping using supination or pronation of the forearm more often resulted in effective trial completions. In contrast, four participants employed a different strategy, requiring the crossing of the DH in front of the NDH in the *Single* D+D setting, which led to hand-tracking loss caused by overlapping hands.

A set of individuals were observed adjusting their strategy upon failure in the indirect cases. Amongst these individuals, the less experienced participants would attempt to use the indirect techniques in a direct manner, returning to familiar patterns. Six participants remarked that D+I felt similar in use to D+D, with no other pairing receiving the same mention. Six participants noted that our system does not lead to awareness of the eyes as an input modality.

6 DISCUSSION

In the following, we discuss our main insights.

6.1 RQ1: Gaze-assisted Hands-on vs. Hands-off

With gaze being capable of ballistic motion, it makes sense that the Acquisition Time for I+I, essentially consisting of just Gaze+Pinch selection, would be lower. Conversely, the lower Preparation Time of D+D and D+I suggests that rotation is easier when directly interacting versus indirectly, which is supported by user feedback, and aligns with studies on multisensory perception in neuroscience.

Interestingly, I+D was most affected, as coordinating the challenging indirect rotation with direct DH input proved difficult. Users often tried to compensate with the DH for the NDH's rotation issues, which impacted DH performance. In contrast, indirect DH input (I+I) allowed better compensation since it avoided direct contact between the object and the hand. These points indicate that D+I should work best, with no NDH rotation issues yet benefitting from less effort and occlusion of the DH. However, this is inconclusive and demands future study; we can see a trend of improved task time and effort compared to other conditions, especially for the more realistic *Continuous* task, but significant results were mainly revealed in contrast to the poorly performing I+D condition.

That said, our results show additional effort as recorded by hand movement when the hand was directly interacting (NDH for D+D and D+I, and DH for D+D and I+D), suggesting a trade-off between ease of interaction and physical energy expenditure.

6.2 RQ2: Sensory-motor Dependencies of D+I vs. I+D

While D+I and I+D seem similar, the difference in which interaction method was used for setting the reference frame (i.e. acquisition and rotation) seemed important to our participants. They ranked D+I better than I+D, remarking a dissatisfaction or difficulty in movement patterns afforded by I+D while praising D+I. Most participants found that D+I offered the most comfort of direct rotation while allowing for manipulation to happen at a comfortable distance. In contrast, seven participants complained that it felt as if I+D enforced, or required, them to do manipulation at a larger distance between their arms, diminishing their comfort and precision.

As such we believe that visual body anchoring of the frame of reference on the NDH is critical for bimanual interactions. The user can benefit from the multisensory integration of having their hand directly touch the object [35], despite the lack of haptics. Indeed, rotation of objects is a complex task that, when performed by the NDH, relies on automatic motor control mechanisms. This may require stronger bottom-up input (and afferent sensory paths) for planning and target acquisition for grabbing an object to make the most of the rotational range of the hand, which is more difficult when the hands are not on the object [24, 26]. Interestingly, the I+D problem was not aggravated with fully indirect.

6.3 RQ3: Dominant Hand also Dominates on Mixed Interaction

While we expected D+I and I+D to offer a balance between performance and usability, compared to I+I, our results showed that I+I outperformed either D+I or I+D on almost all measures and preferences (Acquisition Time, NDH-, DH-, and/or Combined Hand Movement, and Rankings).

Clearly, mixed interaction for bimanual models presented an added complexity for participants. It was not only that participants had to learn a "new" technique but also that this interaction was only applicable to one hand. Normally, we expect symmetry in interaction models and the dominant hand was interfering with the subconscious activities that would have normally controlled the non-dominant hand.

However, this preference for interaction consistency might also suffer from a hidden learning curve, as there are plenty of tasks we perform not only asymmetrically but also out-of-phase, like playing instruments, or more simply, cutting with a knife and fork.

6.4 RQ4: Preshaping as Key Manipulation Pattern

When switching between direct and indirect interaction techniques after finishing a condition, participants became aware of a difference felt in the NDH. The need for clutching and a perceived lack of control during preparation came with switching to the indirect NDH or, inversely, the alleviation of these issues when switching to a direct NDH. We suspect that these perceived and observed differences arise due to deviations from natural reach and grasp behaviour [27]. Our proposed indirect techniques present a departure in the visuomotorics and kinesthetics involved in physical grabbing, impacting the users' mental model and preshaping [18].

Such *preshaping* is a natural human pattern, referring to the anticipatory configuration of the wrist, hand, and fingers to match the size, shape, and orientation of an object, prior to and upon object contact [47]. Preparatory modulation of joint angles in the involved segments of the upper limb is task-driven and gradually evolves through a predictive and responsive phase, typically in parallel with moving the hand to the target [36]. It is plausible that the altering of preshaping behaviour, as seen in I+D and I+I, impacts all aspects of object interaction, as mental rotation, reach-to-grasp, and finally dexterous manipulation are strongly integrated for object-centric goal-oriented actions [18].

6.5 Limitations

We focused on a specific, but common asymmetric bimanual interaction case in this study. Future work could explore further variations to assess the effect of handedness, gaze, and task complexity. Further, future studies can explore other object scales and non-uniform, asymmetric shapes that can offer different rotation affordances. Our framework supports a key feature for users, to transition between the four input modes, through a specific way. Our automatic switching approach may benefit users who frequently change hands or interaction modes (direct or indirect). However, factors such as Midas Touch, context awareness, input transitions [50], and eye-hand coordination [29] could impact the user and are therefore crucial to consider. Alternatively, user-initiated switching, such as through a menu, may be more suitable for those working in specific modelling modes, similar to those in modelling applications [7, 28, 29, 32, 33]. Zooming further out, here we focus on employing the eyes for selection tasks only. Considering the complexity of 3D manipulations, the eyes may provide additional support to the user across various interaction tasks.

7 CONCLUSION

This paper investigated Hands-on, Hands-off interaction as combinations of direct and indirect input for bimanual near-space interaction in 3D where either hand, or both hands, can interact directly or indirectly. We presented five applications showcasing the utility of the interaction combinations for varying purposes, such as intuitively rotating objects directly, while quickly selecting parts indirectly or grabbing an object indirectly to mitigate occlusion while precisely drawing on the surface of the object directly.

In a user study, we focused on understanding bimanual-asymmetric interaction where the NDH rotates and prepares an object as reference for the DH to manipulate parts within. We found that indirect interaction was effective in near-space interaction for manipulating parts of the object and for setting the reference when both hands indirectly interact. However, our users had difficulty effectively using their NDH to indirectly set the reference when the DH directly manipulated parts. During the study, we observed that users would not preemptively rotate their NDH before acquiring the object indirectly, known as preshaping, which we believe to be a major factor.

Overall, this work demonstrates the potential of bringing indirect interaction into the near space in bimanual settings to support interacting with quicker acquisition and lower physical effort, while direct interaction remains preferred for tasks that require a greater sense of intuition.

ACKNOWLEDGMENTS

This work has received funding from the Innovation Fund Denmark (IFD grant no. 9090-00002B) for the MADE FAST project — part of the Manufacturing Academy of Denmark (MADE), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 101021229 GEMINI), and the Danish National Research Foundation under the Pioneer Centre for AI in Denmark (DNRF grant P1).

REFERENCES

[1] Samantha Aziz, Dillon J Lohr, Lee Friedman, and Oleg Komogortsev. 2024. Evaluation of Eye Tracking Signal Quality for Virtual Reality Applications: A Case Study in the Meta Quest Pro. arXiv:2403.07210 [cs.HC]

[2] Ravin Balakrishnan and Gordon Kurtenbach. 1999. Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 56–62. https://doi.org/10.1145/302979.302991

[3] W. Buxton and B. Myers. 1986. A Study in Two-Handed Input. *SIGCHI Bull.* 17, 4 (April 1986), 321–326. https://doi.org/10.1145/22339.22390

[4] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2527–2530. https://doi.org/10.1145/2207676.2208639

[5] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752

[6] Lawrence D. Cutler, Bernd Fröhlich, and Pat Hanrahan. 1997. Two-handed direct manipulation on the responsive workbench. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/253284.253315

[7] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376628

[8] Scott Frees and G Drew Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE, Institute for Electrical and Electronics Engineers, New York, NY, USA, 99–106.

[9] Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.

[10] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1 1988), 139–183. Issue C. https://doi.org/10.1016/S0166-4115(08)62386-9

[11] Ken Hinckley. 2007. Input technologies and techniques. In *The human-computer interaction handbook*. CRC Press, Boca Raton, Florida, United States, 187–202.

[12] Ken Hinckley, Mary Czerwinski, and Mike Sinclair. 1998. Interaction and modeling techniques for desktop two-handed input. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, California, USA) *(UIST '98)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/288392.288572

[13] Ken Hinckley, Randy Pausch, and Dennis Proffitt. 1997. Attention and visual feedback: the bimanual frame of reference. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 121–ff. https://doi.org/10.1145/253284.253318

[14] Ken Hinckley, Randy Pausch, Dennis Proffitt, and Neal F. Kassell. 1998. Two-handed virtual manipulation. *ACM Trans. Comput.-Hum. Interact.* 5, 3 (sep 1998), 260–302. https://doi.org/10.1145/292834.292849

[15] Ken Hinckley, Randy Pausch, Dennis Proffitt, James Patten, and Neal Kassell. 1997. Cooperative bimanual action. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 27–34. https://doi.org/10.1145/258549.258571

[16] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[17] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. https://doi.org/10.1145/123078.128728

[18] Marc Jeannerod. 1999. Visuomotor channels: Their integration in goal-directed prehension. *Human Movement Science* 18, 2-3 (1999), 201–218.

[19] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[20] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300301

[21] Andrea Leganchuk, Shumin Zhai, and William Buxton. 1998. Manual and Cognitive Benefits of Two-handed Input: An Experimental Study. *ACM Trans. Comput.-Hum. Interact.* 5, 4 (Dec. 1998), 326–359. https://doi.org/10.1145/300520.300522

[22] Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Sloth Grønbæk, and Hans Gellersen. 2022. Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 141 (may 2022), 16 pages.

https://doi.org/10.1145/3530882

[23] Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality, In Proceedings of the 2022 Symposium on Eye Tracking Research & Applications. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145, 18 pages. https://doi.org/10.1145/3530886

[24] Antonella Maselli, Eyal Ofek, Brian Cohn, Ken Hinckley, and Mar Gonzalez-Franco. 2023. Enhanced efficiency in visually guided online motor control for actions redirected towards the body midline. *Philosophical Transactions of the Royal Society B* 378, 1869 (2023), 20210453.

[25] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. *Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality.* Association for Computing Machinery, New York, NY, USA, Chapter 15, 7. https://doi.org/10.1145/3448018.3457998

[26] Gonçalo Padrao, Mar Gonzalez-Franco, Maria V Sanchez-Vives, Mel Slater, and Antoni Rodriguez-Fornells. 2016. Violating body movement semantics: Neural signatures of self-generated and external-generated errors. *Neuroimage* 124 (2016), 147–156.

[27] Yves Paulignan and Marc Jeannerod. 1996. Prehension movements: the visuomotor channels hypothesis revisited. In *Hand and brain*. Elsevier, Amsterdam, Netherlands, 265–282.

[28] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[29] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

[30] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. Partially-indirect Bimanual Input with Gaze, Pen, and Touch for Pan, Zoom, and Ink Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2845–2856. https://doi.org/10.1145/2858036.2858201

[31] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

[32] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[33] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-Enhanced Menus in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3385956.3418962

[34] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. 1998. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum* 17 (1998), 12 pages.

[35] Muhtasim Ul Karim Sadaf, Najam U Sakib, Andrew Pannone, Harikrishnan Ravichandran, and Saptarshi Das. 2023. A bio-inspired visuotactile neuron for multisensory integration. *Nature communications* 14, 1 (2023), 5729.

[36] Luis F Schettino, Sergei V Adamovich, and Howard Poizner. 2003. Effects of object shape and visual feedback on hand configuration during grasping. *Experimental Brain Research* 151 (2003), 158–166.

[37] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-Supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2207676.2208709

[38] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300243

[39] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[40] Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) *(VRST '09)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/1643928.1643942

[41] Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In *Human-Computer Interaction – INTERACT 2015*. Springer-Verlag, Berlin, Heidelberg, 315–330. https://doi.org/10.1007/978-3-319-22668-2_25

[42] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (, Hamburg, Germany,) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 252, 15 pages. https://doi.org/10.1145/3544548.3581423

[43] Jia Wang and Robert W. Lindeman. 2015. Object impersonation: Towards effective interaction in tablet- and HMD-based hybrid virtual environments. In *2015 IEEE Virtual Reality (VR)*. Institute for Electrical and Electronics Engineers, New York, NY, USA, 111–118. https://doi.org/10.1109/VR.2015.7223332

[44] Robert Wang, Sylvain Paris, and Jovan Popović. 2011. 6D hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 549–558. https://doi.org/10.1145/2047196.2047269

[45] Shu Wei, Desmond Bloemers, and Aitor Rovira. 2023. A Preliminary Study of the Eye Tracker in the Meta Quest Pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (Nantes, France) *(IMX '23)*. Association for Computing Machinery, New York, NY, USA, 216–221. https://doi.org/10.1145/3573381.3596467

[46] Curtis Wilkes and Doug A. Bowman. 2008. Advantages of velocity-based scaling for distant 3D manipulation. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (Bordeaux, France) *(VRST '08)*. Association for Computing Machinery, New York, NY, USA, 23–29. https://doi.org/10.1145/1450579.1450585

[47] Sara Winges, Douglas Weber, and Marco Santello. 2003. The role of vision on hand preshaping during reach to grasp. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* 152 (11 2003), 489–98. https://doi.org/10.1007/s00221-003-1571-9

[48] Momona Yamagami, Sasa Junuzovic, Mar Gonzalez-Franco, Eyal Ofek, Edward Cutrell, John R Porter, Andrew D Wilson, and Martez E Mott. 2022. Two-In-One: A design space for mapping unimanual input into bimanual interactions in VR for users with limited movement. *ACM Transactions on Accessible Computing (TACCESS)* 15, 3 (2022), 1–25.

[49] 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* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 889–899. https://doi.org/10.1145/3242587.3242643

[50] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. *Gaze-Supported 3D Object Manipulation in Virtual Reality.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445343