Comparing eye–hand coordination between controller-mediated virtual reality, and a real-world object interaction task

Ewen Lavoie
Faculty of Kinesiology, Sport, and Recreation, Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada

Jacqueline S. Hebert
Division of Physical Medicine and Rehabilitation, Department of Biomedical Engineering, University of Alberta, Edmonton, AB, Canada
Glenrose Rehabiliation Hospital, Alberta Health Services, Edmonton, AB, Canada

Craig S. Chapman
Faculty of Kinesiology, Sport, and Recreation, Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada

Virtual reality (VR) technology has advanced significantly in recent years, with many potential applications. However, it is unclear how well VR simulations mimic real-world experiences, particularly in terms of eye–hand coordination. This study compares eye–hand coordination from a previously validated real-world object interaction task to the same task re-created in controller-mediated VR. We recorded eye and body movements and segmented participants' gaze data using the movement data. In the real-world condition, participants wore a head-mounted eye tracker and motion capture markers and moved a pasta box into and out of a set of shelves. In the VR condition, participants wore a VR headset and moved a virtual box using handheld controllers. Unsurprisingly, VR participants took longer to complete the task. Before picking up or dropping off the box, participants in the real world visually fixated the box about half a second before their hand arrived at the area of action. This 500-ms minimum fixation time before the hand arrived was preserved in VR. Real-world participants disengaged their eyes from the box almost immediately after their hand initiated or terminated the interaction, but VR participants stayed fixated on the box for much longer after it was picked up or dropped off. We speculate that the limited haptic feedback during object interactions in VR forces users to maintain visual fixation on objects longer than in the real world, altering eye–hand coordination. These findings suggest that current VR technology does not replicate real-world experience in terms of eye–hand coordination.

Introduction

If you don't own a virtual reality (VR) headset, there's a good chance you know someone who does, showing just how ubiquitous this technology is becoming. In most cases, people use VR for gaming and entertainment as it allows creators to immerse users in fantastical virtual environments. However, the exact same tools can simulate the real world, and experiences can be designed with the desired goal of changing real-world behavior. It is this more practical use of VR that we are interested in studying. Here we focus on the possible utility of VR for improving movements in the real world as with skill training (Lerner, Mohr, Schild, Göring, & Luiz, 2020), medical simulations (Pottle, 2019), sport performance (Oagaz, Schoun, & Choi, 2022), and rehabilitation (Levac, Huber, & Sternad, 2019). In these use-cases, a user can practice a skill in VR that may be high risk in the real world. For example, firefighter trainees can learn to assess the inside of a burning building in VR hundreds of times before ever having to step anywhere near a dangerous real fire. Moreover, VR allows users to practice skills in situations that have a low probability of occurring in the real world (RW) but may have severe negative consequences if performed inappropriately. For example, a surgeon can practice a difficult, emergency surgery in VR that has a low probability of occurring in the real world, reducing the

Citation: Lavoie, E., Hebert, J. S., & Chapman, C. S. (2024). Comparing eye–hand coordination between controller-mediated virtual reality, and a real-world object interaction task. *Journal of Vision*, *24*(2):9, 1–18, https://doi.org/10.1167/jov.24.2.9.

https://doi.org/10.1167/jov.24.2.9 Received May 23, 2023; published February 23, 2024 ISSN 1534-7362

risk of making a mistake when it does occur. Of course, many differences still exist between the real world and simulations using VR headsets, and because of those differences, it's important to explore how closely our behavior in virtual environments translates to the real world. If doing a task in VR leads to real-world behaviors that are suboptimal, are these simulations actually helping? Using the surgical example, what if doing hundreds of hours of simulated VR surgery (Mao et al., 2021) actually leads to worse real-world outcomes because surgeons learned a set of motor plans that don't transfer to a real operating theater?

If we want to compare performance in the real world to performance in a VR environment, especially from a lens of motor skill learning, it's important that we test the right kind of task with the right kind of measures. From this perspective, two important features of people's behavior are how they move in and look at the world. The dance that our eyes and hands engage in when we are manipulating our environment provides a useful tool for uncovering just how immersive and embodying a VR world is (Lavoie & Chapman, 2021). If our eye–hand interactions are the same in VR as in the real world, a strong case can be made that we are behaving as naturally as we do in the real world. Of course, even though technologies like hand-tracking and haptic feedback devices are emerging, in the vast majority of VR deployments, interactions are mediated through handheld controllers. At some level, it might therefore seem obvious that there will be a departure between eye–hand and eye–controller coordination. However, it is important to recognize that, as described above, controller-mediated VR is already being deployed as a proxy for real motor skill learning. So, even though we might scientifically expect differences, commercially this technology is being used in a way that assumes a fundamental—and, importantly, untested—assumption of similarity.

Thus, in the current study, we compare a detailed analysis of eye–hand coordination during an object interaction task performed in the real world and the identical task performed in VR with handheld controllers. We acknowledge that our results might therefore be specific to the task we selected and the VR device we used. But, we hope to make the case that measuring eye–hand coordination during object interactions provides—as a class of tasks—diagnostic power in multiple domains, including the testing of new VR devices. We also acknowledge that object interactions in controller-mediated VR lack the salient haptic feedback that people receive from objects in the real world. As we explain later in this Introduction, without reliable haptic feedback, it is expected that vision will be relied on to ensure successful interactions. But, what we do not know is exactly how the distribution of gaze will change across conditions. For example, one might predict that controller-mediated VR will demand additional gaze resources throughout all phases of a movement or that they are only required when targeting an object, picking it up, or releasing it. Our hope is to show the value of adopting eye and body movement measures as tools to measure behavioral proficiency in virtual environments. This is imperative not only to provide guidance on how VR is being used for real-world skill learning right now but also to set a benchmark against which to compare future human performance in altered realities. As we embrace new technologies that augment or replace our experiences (e.g., mixed-reality headsets, haptic feedback devices, advanced prosthetic limbs), it is important that we have a tool to sensitively document the impact of their adoption.

Eye–hand coordination in the real world

Of course, there are many measures of behavior we could collect in VR and compare to the real world. We use eye–hand coordination during an object interaction task because it is highly stereotypical among humans, and it is traceable back along the primate evolutionary path (Cisek, 2022; Heldstab et al., 2016). It is also fundamental for so many of our daily activities that it is likely tightly linked to the way visual information flows through the brain to control actions. Specifically, it's been shown that the unconscious, dorsal visual stream plays an important role in creating movement plans of hands and arms during object interactions (Desmurget, 1998; Milner & Goodale, 2006) and that the eyes fixate on upcoming targets of pointing movements and obstacles during object manipulation tasks (Johansson, Westling, Bäckström, & Randall Flanagan, 2001; Neggers & Bekkering, 2000). A recent review article explores functional eye and hand movements, highlighting behavioral, neurophysiological, and clinical studies (de Brouwer, Flanagan, & Spering, 2021).

With the ultimate goal of the current study being an in-depth comparison of naturalistic eye–hand coordination between VR and the real world, we discuss eye and hand movements from object interaction studies that allow participants to move their bodies mostly unimpeded. In a previous study, we confirmed that eye movements precede and predict hand reaches for nearly all object interactions and provided evidence that a minimum of approximately half a second of visual fixation on an object before arrival of the hand at that object is necessary for successful computation of grasp dynamics (Lavoie et al., 2018). This "just-in-time" phenomenon is described as the eyes fixating on an object or area immediately prior to its use (Ballard, Hayhoe, & Pelz, 1995; Hayhoe & Ballard, 2005) and, with some contextual flexibility, is quite

consistent, having been found in many environments and populations, including on screens with cursor interactions (Bertrand & Chapman, 2023), upper-limb prosthesis-using populations (Hebert et al., 2019), and even in nonhuman primates (Ngo et al., 2022).

After starting to interact with an object, gaze patterns in the real world retain a consistent pattern with participants disengaging their eyes from an object almost immediately after their hand began or completed interacting with it, making a saccade to the next location of hand action (Land & Hayhoe, 2001; Lavoie et al., 2018). Taken together, this allows us to generate the following summary of real-world eye–hand coordination during naturalistic, sequential object interactions:

1. When the goal is to pick up and move an object, people need to visually fixate approximately 500 ms before the hand arrives at the object. This duration can be altered by the distance of the object. For example, this time could be extended if it's going to take the hand a long time to get to the object (e.g., object is across the room) or truncated if the objects are closer together (Lavoie et al., 2018).
2. People will maintain visual fixation on the object they are picking up until they are confident that they will be successful in initiating a movement. This can be nearly instantaneous after the hand has arrived at the object, as with mouse clicks (Bertrand & Chapman, 2023), or can take an extended period of time, as in the case of object movements carried out by prosthesis users (Hebert et al., 2019; Lavoie et al., 2018).
3. Once that specified level of confidence has been achieved, visual fixation will immediately shift from the object to its future drop-off location. The duration of this advanced fixation is tied to the length of the distance the object needs to travel. If the transport distance is short, as with mouse clicks, then the advanced look will be short. But if transport distance is long, as with walking across a kitchen when making tea, then the advanced look will be long.
4. The eyes will remain fixated on a drop-off target for a minimum of half a second and as long as it takes for the object to confidently be released. Even for short, confident transports like mouse movements, the visual fixation lingers at the drop-off location until about 500 ms has been reached (Bertrand & Chapman, 2023). For slow (>500ms), confident transports, like transporting a pasta box in the real world, the eyes instantaneously shift away upon releasing the box (Lavoie et al., 2018).

This generalized understanding of eye–hand coordination during naturalistic object interactions in the real world gives us a measuring stick against which we can compare eye–hand coordination during similar tasks in VR.

Predictions for eye–hand coordination in VR

There are theoretical reasons why humans may fundamentally perceive VR differently. Snow and Culham (2021) break down the key differences between real, tangible objects and several levels of proxy objects, including two-dimensional images, three-dimensional images, and objects simulated as real through VR. This work shows that humans recognize real objects better than proxies, remember real objects better than proxies, and have greater attention toward real objects than proxies. Ultimately, and pertinent to the current study, which uses an object interaction task, this is thought to be driven by the action affordance of real objects, compared to proxies (Marini, Breeding, & Snow, 2019; Snow & Culham, 2021).

Another reason we may treat virtual objects differently than real-world objects is the availability of haptic feedback. In the real world, when we reach out with our limb to interact with an object, we receive proprioceptive feedback and eventually haptic information at our fingertips, like the pressure of our fingers on the object as we secure it and the weight of the object as we lift it. For the same action in controller-mediated VR, we have a much less salient experience. Here, we may feel similar proprioception in our limb but not in our hand as it is grasping a plastic controller. When we are able to initiate a grasp, at most we may receive a slight vibration from the plastic controller—with no information about pressure or object weight. Indeed, this impoverished haptic feedback is cited as one of the reasons we may shift from using the more dorsally driven, vision-for-action visual information stream, thought to underlie real eye–hand coordination, to using the vision-for-perception ventral visual stream (Harris, Buckingham, Wilson, & Vine, 2019). This hypothesis rests heavily on findings showing that pantomimed movements to remembered objects elicit greater ventral stream engagement than movements to objects in real time (Goodale, Jakobson, & Keillor, 1994). The consequences of a ventral stream shift to eye–hand coordination are largely unknown.

If the lack of haptic feedback is predicted to drive differences in movement and gaze, what specific alterations in eye–hand coordination might we observe? Here we can turn to work with upper-limb prosthesis users as a possible model. Upper-limb prosthesis users also interact with objects while receiving limited haptic feedback, similar to controller-mediated VR users. We know that prosthesis users move much slower, visually fixate for longer on objects before interacting with them, maintain fixation on objects for much longer after interacting with them, and spend

much more time fixating on their own limb while performing object interaction tasks (Hebert et al., 2019). Even more interesting, when touch sensation is returned to prosthesis users through a combination of reinnervation surgery, grip kinesthesia, and intuitive motor control, their behavioral patterns shift closer to those of the able-bodied population (Marasco et al., 2021). Most notably, the return of some haptic feedback liberates the eye gaze to move away from dedicated fixations to the end effector and make more advance fixations toward upcoming movement targets (Hebert & Shehata, 2022).

Taken together, there is good reason to predict differences in gaze distribution between real and virtual object interactions. In this study, we provide a test of this prediction and, in doing so, bring a specificity to the visuomotor processes engaged—not just *that* eye–hand coordination is different but *how.* To the crux of the problem we wish to investigate—if VR elicits visuomotor behaviors that are not well matched to the real world, we argue they may lack utility as a skill learning tool. Motor skill learning is highly contextual, so learning to perform a task in VR that elicits a known difference in visuomotor strategy might actually be counterproductive. Moreover, we hope that by comparing detailed eye–hand coordination patterns in VR to a previously validated object interaction task from the real world (Boser et al., 2018; Hebert et al., 2019; Lavoie et al., 2018; Valevicius et al., 2018, 2019; Williams, Chapman, Pilarski, Vette, & Hebert, 2019), we expose its utility as a benchmark for future advancements in the assessment of human motor performance.

Methods

The data used here are a reanalysis and comparison of data from two previous studies. The first data set has been used to publish an exploration of eye-movement patterns during object interactions in the real world (Lavoie et al., 2018), while the second compares embodiment and movement differences between two virtual reality limb visualizations (Lavoie & Chapman, 2021). Here we describe the details necessary for our comparison of parts of each of these data sets and, for full details, encourage readers to seek out those previous publications.

Participants

Real world

This previously published study contained 24 able-bodied adults, with no upper-body pathology or history of neurological or musculoskeletal injuries within the past 2 years (Boser et al., 2018; Lavoie et al., 2018; Valevicius et al., 2018, 2019). Participants provided written informed consent to participate in the study. Seven participants were dropped due to software and hardware issues, including insufficient ability to track eye movements, freezing of collection computer, and missing motion-capture markers. The remaining 17 participants (9 male, 8 female) had an average height of 173 ± 10.9 cm and were made up of 16 self-reported preferred right-hand users and 1 self-reported preferred left-hand user. All participants had normal or corrected-to-normal vision, while 2 participants were tested without corrected vision, as they removed their glasses to don the eye tracker. These participants assured the experimenters they could complete the task normally. All participants were unaware of the purposes of the experiments. All procedures were approved by the University of Alberta Health Research Ethics Board (Pro00054011), the Department of the Navy Human Research Protection Program, and the SSC-Pacific Human Research Protection Office.

Virtual reality

Twenty-one self-reported right-handed undergraduate students received course credit and provided informed consent to participate in this study. Fourteen participants (11 male, 3 female) with an average height of 171.1 ± 8.87 cm from this VR condition were used for this study, with 7 participants being dropped (6 due to poor eye-tracking data and 1 due to a software issue during data collection). Eight of the remaining 14 participants removed their glasses, and 1 participant was color blind and was told the colors of the placement targets by the experimenter. Procedures were approved by the University of Alberta Health Research Ethics Board (Pro00085257).

Apparatus

Real world

Participants were fitted with a head-mounted, binocular eye tracker (Dikablis Professional 2.0; Ergoneers GmbH, Manching, Germany). They were asked to position the headset comfortably before experimenters tightened the built-in elastic strap on the back to hold it steadily in place. In addition to the head-mounted eye tracker, 57 upper-body motion-capture markers were placed on the participant and were tracked with 12 infrared cameras (Bonita; Vicon Motion Systems, Oxford, UK), including markers on the index finger and thumb and a plate with three markers on the back of the hand. Additional markers were placed

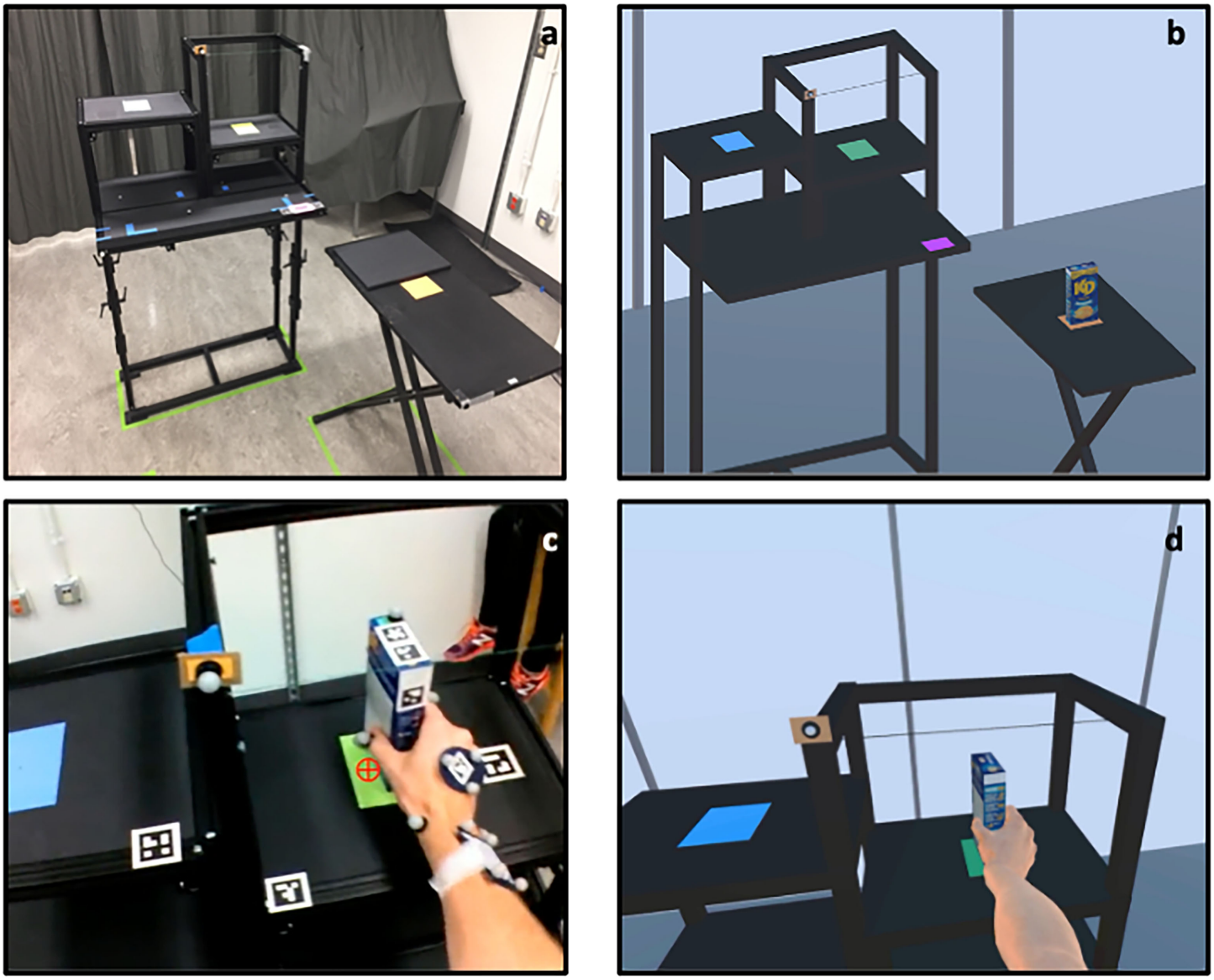

Figure 1. The apparatus of the Pasta Box Task in the real world (**a**) and (**b**) in VR. The first-person view of a participant carrying out the task in the real world (**c**) and (**d**) in VR.

on the pasta box and other task-relevant parts of the apparatus. The Pasta Box Task was developed as part of the DARPA HAPTIX project, as an assessment tool for upper-limb prosthesis users (Boser et al., 2018; Hebert et al., 2019; Lavoie et al., 2018; Valevicius et al., 2018, Valevicius et al., 2019). The apparatus consists of a shelving unit, with an accompanying side table (Figure 1a). A first-person view of a participant performing part of the task can be seen in Figure 1c.

Virtual reality

Participants donned an HTC Vive head-mounted display (HMD, Vive; HTC and Valve, New Taipei City, Taiwan, and Bellevue, WA, USA, respectively) with a Deluxe Audio Strap and inserted binocular eye trackers collecting pupil position at 200 Hz (PupilLabs GmbH, Berlin, Germany). They were immersed in a model of our lab space (built in Unity [Unity Technologies, San Francisco, CA, USA] and using NewtonVR [Today Tomorrow Labs, Seattle, WA, USA]). The virtual task apparatus was built to the same measurements and appearance as the real-world task described above, consisting of a set of shelves with three placement targets and a pasta box (Figure 1b). Participants held an HTC Vive controller in each hand for the duration they wore the HMD, while what they saw was a set of dynamic limbs (Figure 1d, and Supplementary video).

Procedure

Real world

The data used here were collected as part of a larger experiment. Each participant underwent six data collection blocks: three of the pasta task that we used here and three of another object movement task, which we did not use in this study. Only one of these three pasta task blocks was used here as it is the only block where both eye and body movements were collected. The sets of collection blocks were randomized in order for each participant. Before each collection block, participants underwent a series of calibration exercises, fully described in Lavoie et al. (2018). Each collection block consisted of as many trials as necessary to obtain at least 20 trials without errors.

Participants performed the object movement task with their right hand. The task was designed to assess the coordination of gaze and movement during everyday object interactions. Each trial was initiated with an auditory cue and consisted of three object interactions. First, participants moved the pasta box from the Start/End Target on a table on their right side onto the Mid Shelf Target in front of them (Figure 2a). Then, participants move the pasta box from the Mid Shelf Target to the High Shelf Target by crossing the body's midline (Figure 2b). Finally, the pasta box was picked up from the High Shelf Target and placed back on the Start/End Target (Figure 2c). At the start and end of each trial and after each pasta box placement, participants touched the Home position (pink rectangle in Figure 2a–c) and were instructed to visually fixate on a small gray sphere (neutral position) before each trial began and after each trial was completed.

Participants were instructed to move at a comfortable pace and interact with the pasta box on its side. There were colored targets indicating where the pasta box should be placed for each movement, and participants were instructed to place the box on the short edge within the boundaries of each placement target. Additionally, participants were to avoid dropping the pasta box, contacting the apparatus, hesitating, or making undesired movements (like scratching one's leg). If a rule was violated, participants were told to complete the trial to the best of their ability and an extra trial was added at the end of that group of trials.

Virtual reality

After donning the VR headset (HTC Vive) but before entering the VR lab environment, participants carried out a brief (∼15 s) eye-tracking calibration (PupilLabs GmbH, Berlin, Germany), prompting them to fixate one-by-one on a set of small gray circles presented virtually in a larger circle in front of them. The experiment consisted of two counterbalanced sessions of at least 20 error-free repetitions of the object interaction task. One session showed participants virtual models of the plastic controllers they were holding, while in the other session, participants saw a virtual representation of arms that extended from their torso, with hands that moved spatially with the plastic controllers they held (Full Arms VR [Bad Plan Games]). For this study, only the VR condition with a virtual representation of arms was used.

Participants used the plastic controller in their right hand to interact with the virtual pasta box. This interaction was governed by a 5 -m diameter invisible sphere with its center located approximately 10 cm distal to the participant's real-world hand. The plastic controller vibrated when this sphere intersected the pasta box or Home position. Vibration indicated the participant could initiate an interaction with the pasta box by pulling the trigger button. When the trigger was depressed >50%, an interaction began and the pasta box moved with the plastic controller until the trigger was released (<50%).

Data processing

Real world

Custom software was used to trigger the collection of the eye- and motion-tracking software simultaneously and synchronize these data streams for segmentation and analysis. With our custom GaMA software, we used the x and y coordinates of each eye from the video frame of the eye tracker and combined it with the calibration data, including the movement of the head and objects in a regression function, generating a virtual location of the participant's gaze (as represented by a gaze vector) in the coordinate frame of the motion-tracked objects and body.

Using a combination of hand and object velocities and positions, each trial was segmented into its three object movements, with each object movement subsequently segmented into Reach (hand moving toward object), Grasp (hand starting the object interaction), Transport (hand moving object between locations), and Release (hand completing the object interaction and moving away) phases (see Lavoie et al., 2018, for full segmentation details and Figure 3 for sample segmentation). A fifth Home phase (when the hand was returning home after completing each movement) was also segmented from the data but not included in the analysis.

Because we are interested in overt fixations to areas relevant to object interactions, and since previous research has shown that participants rarely fixate on objects or areas irrelevant to the goal of a task (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, 2009; Land

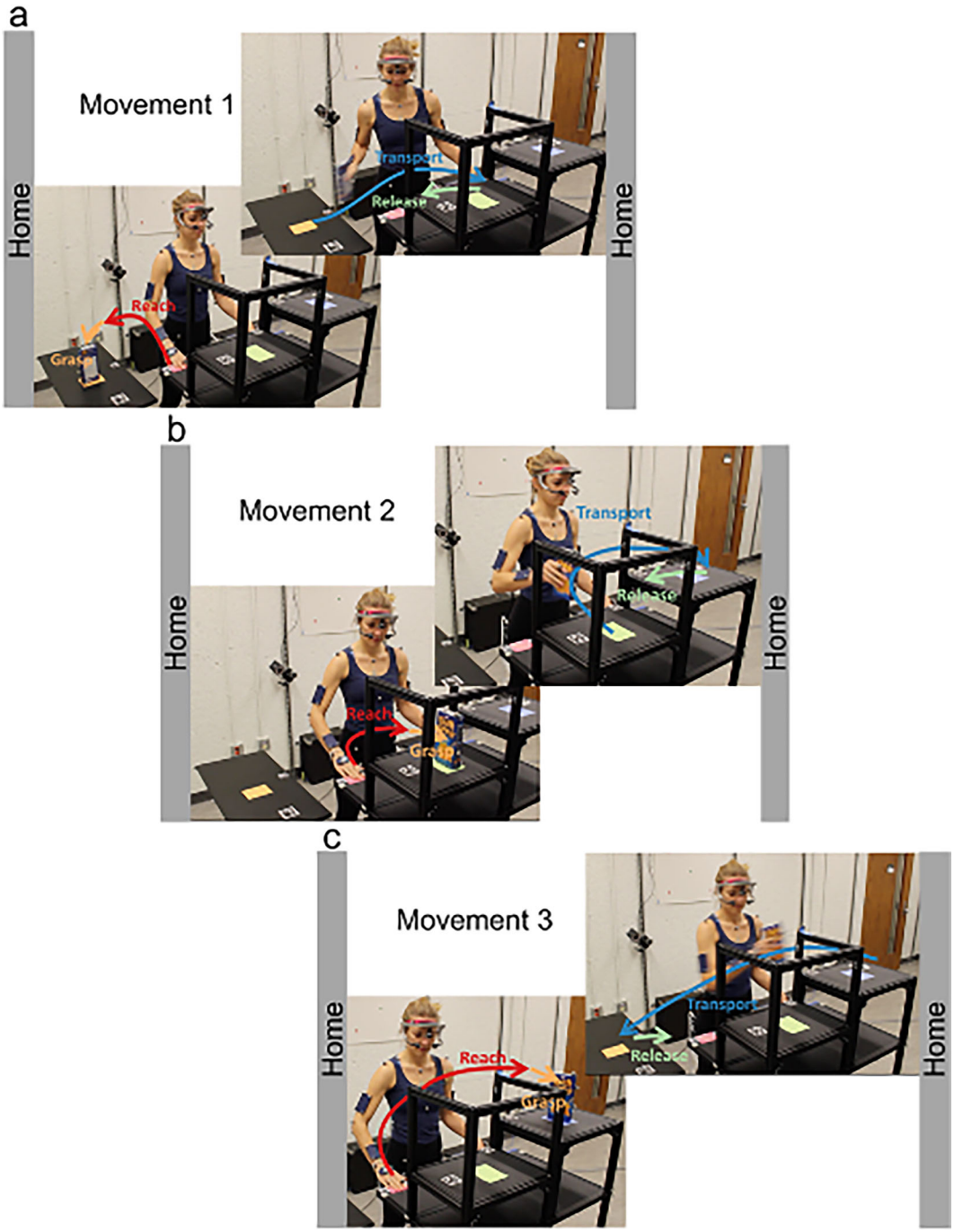

Figure 2. The Pasta Box Task includes Reach, Grasp, Transport, and Release of a pasta box at three target locations. (**a**) Movement 1: Grasp from side cart (Start/End Target) and Release on Mid Shelf Target. (**b**) Movement 2: Grasp from Mid Shelf Target and Release on High Shelf Target. (**c**) Movement 3: Grasp on High Shelf Target and Release on Start/End Target (first published in Lavoie et al., 2018).

& Hayhoe, 2001; Lavoie et al., 2018; Tatler, Hayhoe, Land, & Ballard, 2011), we selected specific regions during each phase of movement for analysis. For this study, the areas of interest (AOIs) within each phase were defined as the current location being acted on by the hand (*Current*) and the hand itself or an object being moved by the hand when no other AOI is being fixated (*Hand in Flight*). A fixation to an AOI was said to occur when the distance between gaze vector and AOI was sufficiently small and the velocity from gaze vector to AOI was also sufficiently low. To account for blinks, any brief periods of missing data in each AOI fixation were filled in. Then, to avoid erroneous fixation detection (e.g., fly-throughs), any brief fixations were removed. Full details can be found in a previous publication (Lavoie et al., 2018).

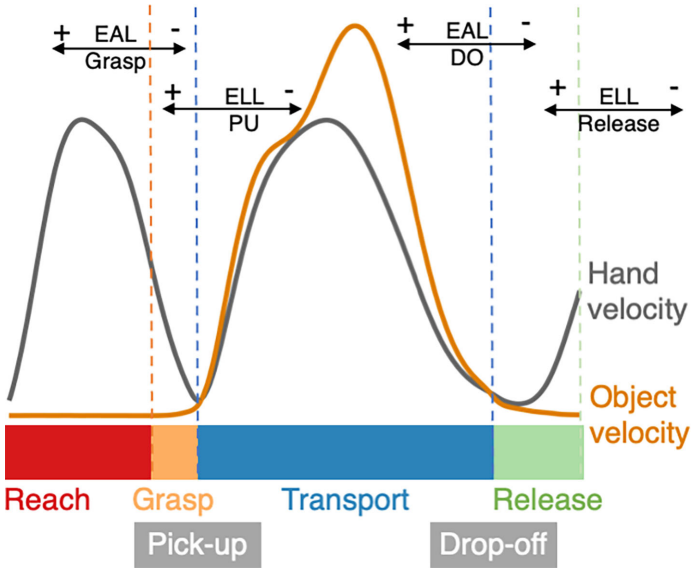

Figure 3. The segmentation of an object Movement into its Reach, Grasp, Transport, and Release phases is determined by the velocity of the object (orange trace), the velocity of the hand (grey trace), and distances to task-relevant locations. Also shown are the approximate temporal locations defined by the terms Pick-up and Drop-off, as well as the Eye Arrival Latency (EAL) and Eye Leaving Latency (ELL) measures associated with each (adapted from Lavoie et al., 2018).

Virtual reality

Using custom C# scripts, the three-dimensional position and rotation of each plastic controller, the HMD (HTC Vive), pasta box, placement targets, and other relevant objects were recorded (90 Hz) on each trial. We then used the exact same segmentation procedure from the real world to generate equivalent events and measures for each trial of our VR data. To further emphasize this point, the real-world and VR raw data were treated identically, with only the following two minor adjustments: In VR, eye and motion data were synchronized during collection through Unity, and in VR, the gaze vector was automatically generated by the PupilLabs software.

Dependent measures and predictions

The real-world measures here are based on those from our previously published study (Lavoie et al., 2018), but values may differ slightly as the segmentation and measure generation procedures are continually being improved. We grouped our measures into three families answering three broad questions. Family 1 includes the *Absolute Duration* and *Relative Duration* of each phase (Reach, Grasp, Transport, Release) and answers the question: "How long do people take?" Family 2 includes the *Number of Fixations* and the *% Fixation Time* to the Current AOI in each phase and to the Hand in Flight AOI during the Reach and Transport phases and answers the question: "Where do people look?" For Family 3, we calculated the *Eye Arrival Latency*, which measures when the gaze lands on a location around the time a Grasp begins and Transport ends (Drop-off), and also the *Eye Leaving Latency*, which measures when the gaze leaves a location around the time when Transport starts (Pick-up) and when Release ends. These four latency measures were calculated for each object movement and answer the question: "When do people look?" The definition and our predictions for each measure are as follows:

Family 1: How long do people take?

- *Absolute Duration*
  - The time in seconds spent in each phase as determined by our segmentation.
  - The sum of the Absolute Durations of each phase equals the Total Movement Time for each of the three Movements.
- *Relative Duration*
  - The percentage of time each phase contributes to the Total Movement Time determined by our segmentation.

We predict that participants will move slower in VR because, for most, VR is a novel medium, and with novelty comes hesitation (Joyner, Vaughn-Cooke, & Benz, 2021). We expect to find greater *Absolute Duration* values in VR for each of the movement phases compared to the real world, but with disproportionately large values in the Grasp and Release phases, the moments when participants are initiating and terminating an object interaction with limited haptic feedback. We predict this will show up in the *Relative Duration* values as well. Although the general proportions of each phase of movement will be more similar than different between VR and the real world, we expect VR to have larger *Relative Duration* values for Grasp and Release phases and, as a result, smaller *Relative Duration* values for Reach and Transport phases. Thus, we expect the lack of haptic feedback in VR will cause participants to move with absolute and relative durations more similar to prosthesis-using populations (Hebert et al., 2019).

Family 2: Where do people look?

- *Number of Fixations*
  - The number of distinct (separated by at least 100 ms), continuous (> 100 ms) fixations to an AOI in a given phase.
- *% Fixation Time*
  - The amount of time fixated on an AOI in a phase divided by the Absolute Duration of that phase, multiplied by 100.

We expect fairly similar patterns of visual fixation between the real world and VR versions of the task. That is, we predict participants will fixate objects and areas that they are interacting with or will be interacting with in the future, with little visual attention to objects and areas irrelevant to the task (Hayhoe & Ballard, 2005; Land & Hayhoe, 2001; Lavoie et al., 2018). However, we anticipate that fixations to a participant's own hand, and the object in their hand, will increase while transporting the object in VR, akin to prosthesis users who also lack haptic feedback (Hebert et al., 2019).

In general, we expect that the *Number of Fixations to Current* and to the *Hand in Flight* will be similar between VR and the real world except during Transport. Here we predict participants may increase their *Number of Fixations to Current* and to the *Hand in Flight* in VR as a result of the need to fixate back and forth on the object in their hand, to be sure that the object is being successfully transported, and back again to the drop-off target.

Keeping the "just-in-time" phenomenon in mind, participants will likely fixate about half a second before their hand arrives at a location to pick up or drop off the box. When coupled with the increase we expect to see for the *Absolute Duration* values, we predict that the *% Fixation Time to Current* during Reach and Transport will be lower in VR than in the real world. Similarly, considering our predicted need for participants in VR to look more toward their hand while it Transports an object, we expect to see a disproportionately low *% Fixation Time to Curren*, and high *% Fixation Time to the Hand in Flight* during Transport while in VR.

Finally, owing to degraded haptic feedback, we think participants will have reduced confidence in the success of the start and end of an object interaction, resulting in *% Fixation Time to Current* during Grasp and Release that will be higher in VR than in the real world. Even though these phases will likely be longer in VR, we believe participants will have to fixate for the majority of each of them to ensure a successful interaction occurs.

Family 3: When do people look?

See Figure 3 for a visual description.

- *Eye Arrival Latency at Grasp (EAL Grasp)*
  - EAL Grasp is defined as Grasp start time minus the time of eye arrival at the Grasp location.
- *Eye Leaving Latency at Pick-up (ELL PU)*
  - ELL PU is defined as Transport start time minus the time of the eye leaving the Pick-up location or object.
- *Eye Arrival Latency at Drop-off (EAL DO)*
  - EAL DO is defined as Transport end time minus the time of the eye arriving at the Drop-off location.
- *Eye Leaving Latency at Release (ELL Release)*
  - Eye Leaving Latency at Release is defined as Release end time minus the time of the eye leaving the Release location or object.

Finally, here we predict the differences between VR and real-world behavior will extend to the temporal dynamics of eye–hand coordination. Importantly, we predict that the "just-in-time" phenomenon will be preserved, as it has been across numerous studies and contexts (Ballard et al., 1995; Bertrand & Chapman, 2023; Hayhoe & Ballard, 2005; Hebert et al., 2019; Land & Hayhoe, 2001; Lavoie et al., 2018; Ngo et al., 2022). That is, in VR, we expect that participants will begin fixating on an object at least half a second before their hand arrives to pick it up, leading to an *EAL Grasp* that will be similar in VR and the real world. In the same vein, we predict participants will fixate the drop-off target of the box for about the same amount of time prior to the box arriving in VR and the real world, giving similar *EAL DO* measures.

However, due to the lack of haptic feedback, we expect VR participants to continue fixating for much longer than real-world participants on the object they are picking up or releasing, as a way to ensure that the start and end of an interaction are successful. Therefore, we predict the *ELL PU* and *ELL Release* values will be much larger in VR compared to the real world. Because participants in VR lack haptic feedback, they will need to use visual attention to take its place to ensure a successful object interaction has occurred. Again, these predictions in VR are based on prosthesis users, who lack detailed touch information while interacting with objects, and so adapt their visual fixation patterns to be able to complete the task successfully (Hebert et al., 2019; Marasco et al., 2021).

Overview of statistical analysis

With the large number of tests and the fact that we were aiming to find ways in which the VR condition differed from the real-world condition, we used a modified procedure developed by Cramer and colleagues to correct for our large number of tests (Cramer et al., 2016). We divided measures into three families of results: Family 1 (How long do people take?), Family 2 (Where do people look?), and Family 3 (When do people look?) and listed the *p*-value (Greenhouse–Geisser corrected if available) of every mixed analysis of variance (RMANOVA) comparing the VR condition to the real-world condition in descending order (most to least significant) within each family. Using the formula, Adjusted $\alpha = [0.05] / [(\#\text{ of tests}) - (\text{rank order} - 1)]$, and selecting the tests whose calculated *p*-values were less than their Adjusted α, we had a conservative list of tests we would move

forward with. Any Omnibus Mixed analysis of variance (ANOVA) that showed significant interaction effects of Condition were followed up with a similar process in which each *p*-value from the post hoc tests for this measure were listed in descending order. Using the same formula as above and the number of tests for this follow-up set, any significant main and interaction effects were found. From this, all possible pairwise comparisons of the relevant factors were conducted using a Bonferroni correction with a corrected $p < 0.05$ marking a significant effect.

It should be repeated that our main empirical interest was to examine behavior across the virtual and real-world conditions. Therefore, any effects that were exclusively due to Movement and/or Phase are not discussed here, and we refer the reader to our previous work for full details (Lavoie et al., 2018).

For each participant in both conditions, each of the dependent measures was calculated for every trial, then averaged across trials. Below is a breakdown of the statistical analyses run within each family of tests.

Family 1: How long do people take?

For both *Absolute Duration* and *Relative Duration*, we ran Condition (RW vs. VR) × Movement × Phase Mixed ANOVAs, where there were two Conditions and three Movements. *Absolute Duration* and *Relative Duration* were split into four phases (Reach, Grasp, Transport, Release).

- Condition (RW vs. VR) × Movement × Phase Mixed ANOVA
  - *Absolute Duration*: two Conditions, three Movements, four Phases
  - *Relative Duration*: two Conditions, three Movements, four Phases

In addition, because of its novelty, we tested whether participants would show a different learning effect in VR and therefore show a steeper decrease in duration in VR compared to the RW. To test this, we compared the change from the first five trials to the last five trials between the two conditions. We calculated the average *Total Movement Time* (sum of the *Absolute Duration* of all five phases, including the Home phase) for the first five and last five trials for each participant in VR and the real world. Each participant had a value for the first five trials and for the last five trials. This was titled the Trial Position. We ran a 2 × 2 Mixed ANOVA of Trial Position × Condition specifically searching for an interaction effect between Trial Position and Condition.

Family 2: Where do people look?

For both *Number of Fixations to Current*, *% Fixation Time to Current*, *Number of Fixations to Hand in Flight*, and *% Fixation Time to Hand in Flight*, we ran Condition (RW vs. VR) × Movement × Phase Mixed ANOVAs. There were two Conditions, three Movements, and four Phases (Reach, Grasp, Transport, Release) for the measures to Current, while there were two Conditions, three Movements, and only two Phases (Reach, Transport) for the measures to the Hand in Flight. This reduction in phases during Grasp and Release is because the participant's hand is at the pick-up or drop-off location and is indistinguishable from the target.

- Condition (RW vs. VR) × Movement × Phase Mixed ANOVA
  - *Number of Fixations to Current*: 2 Conditions, 3 Movements, 4 Phases
  - *% Fixation Time to Current*: 2 Conditions, 3 Movements, 4 Phases
  - *Number of Fixations to Hand in Flight*: 2 Conditions, 3 Movements, 2 Phases
  - *% Fixation Time to Hand in Flight*: 2 Conditions, 3 Movements, 2 Phases

Family 3: When do people look?

To compare the eye latency measures (*EAL Grasp*, *ELL PU*, *EAL DO*, *ELL Release*), we carried out a 2 × 3 Mixed ANOVA of Condition (RW vs. VR) × Movement for each.

- Condition (RW vs. VR) × Movement Mixed ANOVA
  - *EAL Grasp*: two Conditions, three Movements
  - *ELL PU*: two Conditions, three Movements
  - *EAL DO*: two Conditions, three Movements
  - *ELL Release*: two Conditions, three Movements

Results

As the amount of data and possible comparisons are rather large, here we focus on the main similarities and differences found in our behavioral measures between the real world and in VR. In the Methods section, we describe our correction procedure for the large number of statistical tests we did. Here we provide the results after this correction, and only state which tests yield significant differences. Full results and statistical analyses with significance levels will be provided upon request.

Family 1: How long does it take people to move?

Absolute duration in VR twice as long as RW: First, we compared the *Absolute Duration* of each phase of movement between VR and the real world, with

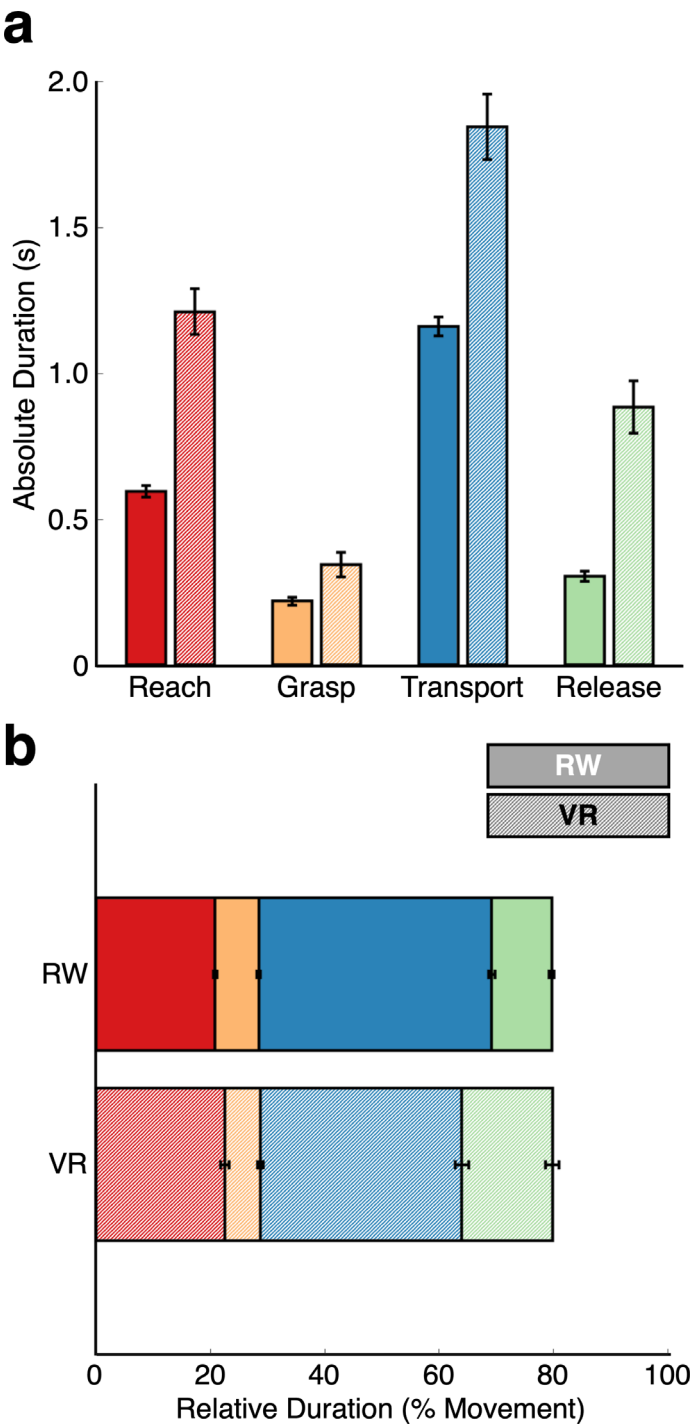

Figure 4. The average of all participants' (**a**) *Absolute Duration* (s) and (**b**) *Relative Duration* (%) of each phase of movement of the Pasta Box Task in VR and RW. Error bars represent standard error of the mean (*SEM*). In both (**a**) and (**b**), the Home phase is omitted as it is irrelevant to the object interaction.

VR being much longer in each (Figure 4a). The most pronounced increase in *Absolute Duration* between the two conditions was found in the Release phase (RW = 0.30 s, $SD = 0.07$ s; VR = 0.88 s, $SD = 0.34$ s; $p = 4.76 \times 10^{-8}$), and the least pronounced increase was found in the Grasp phase (RW = 0.22 s, $SD = 0.06$ s; VR = 0.35 s, $SD = 0.16$ s; $p = 4.00 \times 10^{-3}$).

As predicted, participants in VR took much longer to complete the task than participants in the real world. In fact, the *Total Movement Time* of each trial in VR was approximately twice that of the RW (RW = 8.73 s, $SD = 1.19$ s; VR = 16.51 s, $SD = 3.70$ s; $p = 4.88 \times 10^{-9}$). We compared the change from the first five trials to the last five trials in the two conditions to test for a possible difference in the rate of learning. *Total Movement Time* in both VR (VR: first five = 17.77 s, $SD = 4.71$ s; last five = 15.75 s, $SD = 3.62$ s; $p = 3.90 \times 10^{-2}$) and the real world (RW: first five = 9.01 s, $SD = 1.26$ s; last five = 8.57 s, $SD = 1.17$ s; $p = 2.67 \times 10^{-4}$) did indeed show the predicted speeding up with practice, but there was no significant effect or interaction with condition found, providing no evidence for different learning rates across environments.

Predictable durations of each phase of movement: Overall, people move slower in VR, but our results from *Relative Durations* suggest that the relative distribution of time over the course of an object interaction is mostly the same between conditions (Figure 4b). That is, the most time is spent Transporting and Reaching for an object in both VR and the real world, with less time spent Releasing and even less Grasping. However, despite this general similarity, the fact that the *Absolute Durations* during Grasp were more similar while the *Absolute Durations* during Release were less similar between conditions leads to commensurate differences in the *Relative Duration* measures, with subsequent effects to all phases of movement.

First, participants spend slightly more ($p = 8.00 \times 10^{-3}$) relative time in the Reach phase in VR (22.52%, $SD = 2.84\%$) than in the RW (20.82%, $SD = 4.05\%$) and slightly less relative time in the Grasp phase in VR (RW = 7.64%, $SD = 1.39\%$; VR = 6.25%, $SD = 1.92\%$; $p = 1.90 \times 10^{-2}$). *Relative Duration* is much lower ($p = 5.79 \times 10^{-5}$) in the Transport phase in VR (35.20%, $SD = 4.47\%$) compared to the real world (40.71%, $SD = 2.33\%$) and much higher ($p = 4.20 \times 10^{-5}$) in the Release phase in VR (15.87%, $SD = 4.46\%$) compared to the real world (10.54%, $SD = 1.84\%$). The difference in Release *Relative Duration* is the most pronounced. It should be noted that because this measure is designed to be proportional, when one measure accounts for "more of the pie," another measure takes up "less of the pie." Here, in VR, we see the *Relative Duration* during Grasp take up "less of the pie," resulting in the *Relative Duration* during Reach taking up more. While, again in VR, we see the *Relative Duration* during Release take up much "more of the pie," leaving less for the *Relative Duration* of Transport.

The two major takeaways from this measure are that the distribution of time to phase is similar across VR and the RW except that in VR, there is a significant extension of the time spent releasing.

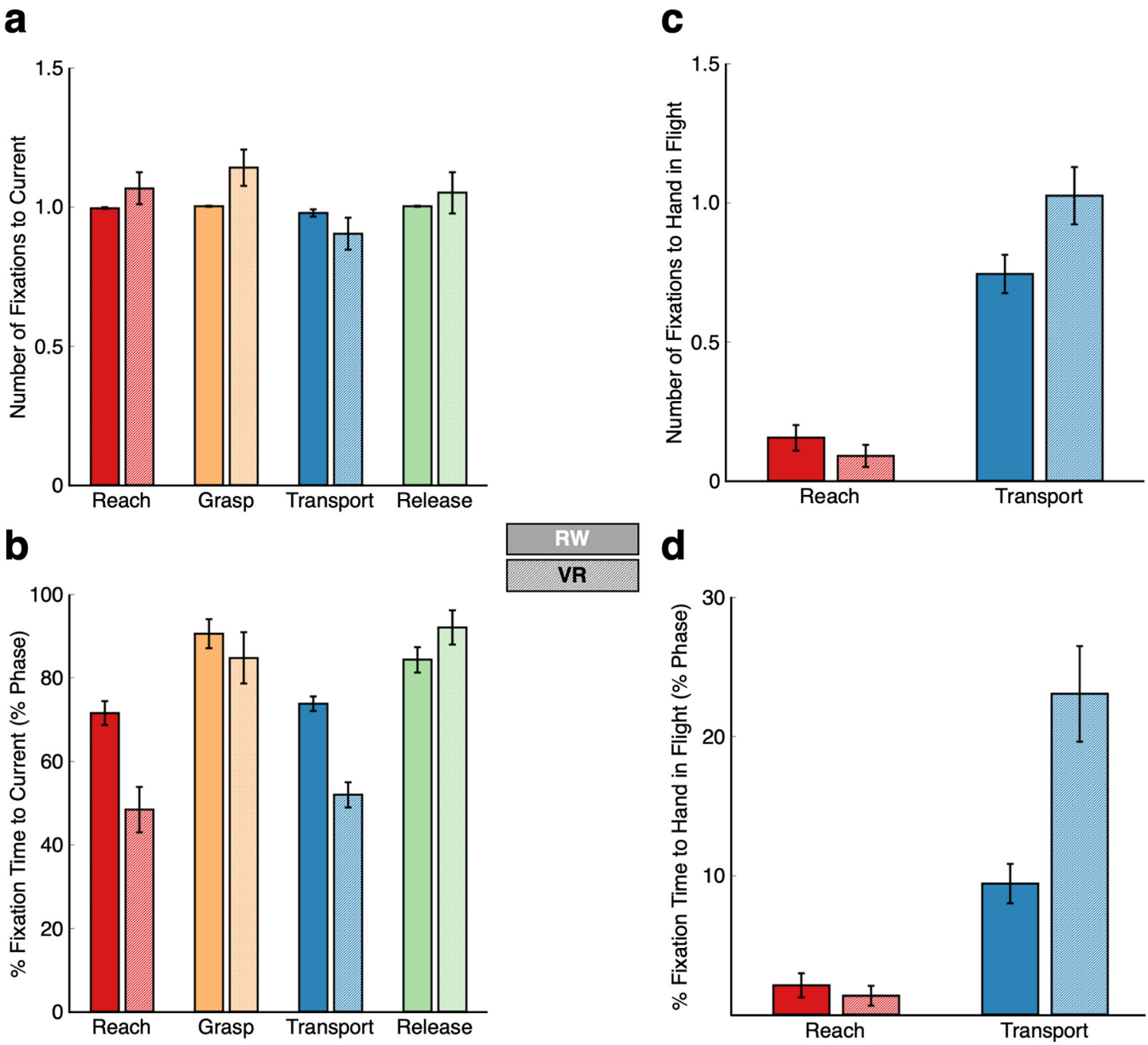

Figure 5. The average of all participants' (**a**) *Number of Fixations to Current* (#), (**b**) *% Fixation Time to Current* (%), (**c**) *Number of Fixations to Hand in Flight* (#), and (**d**) *% Fixation Time to Hand in Flight* (%) of each phase of movement of the Pasta Box Task in VR and RW. Error bars represent *SEM*.

Family 2: Where do people look?

Participants fixate temporally relevant objects and areas: Participants' visual fixation behavior in VR followed the same characteristic pattern as we found in the real world (Lavoie et al., 2018). When planning to move an object from one location to another, participants almost always fixate on the object for approximately half a second before their hand arrives to interact with it. They then stay fixated on the object as they initiate an interaction before breaking this fixation once they're confident the interaction is a success. At this moment, the eyes shift to the drop-off location and fixate there until the hand, along with the object, arrives to release the object. Almost no fixations are made to objects and areas that are irrelevant to the object movement being made. In fact, we found no significant differences in the *Number of Fixations to Current* (Figure 5a) between

the real world and VR. Despite these broad similarities, there are notable differences in the distribution of gaze between the real world and VR, especially when factoring in the total duration differences across the two environments (Figure 5).

During the Reach and Transport phases, participants in the real world consistently spent a larger proportion of time fixating on the location they were about to interact with (*% Fixation Time to Current*, Figure 5b) compared to VR participants (Reach to Box at pickup: RW = 71.56%, *SD* = 11.88%; VR = 48.41%, *SD* = 20.54%; $p = 1.15 \times 10^{-4}$; Transport to drop-off location: RW: 73.86%, *SD* = 7.27%; VR = 52.00%, *SD* = 11.46%; $p = 1.26 \times 10^{-9}$). Conversely, no significant differences between the RW and VR for *% Fixation Time to Current* were found during either the Grasp (RW = 90.63%, *SD* = 14.38%; VR = 84.81%, *SD* = 23.15%) or Release (RW = 84.34%, *SD* = 12.57%; VR = 92.12%, *SD* = 15.43%) phases.

It is important to consider this pattern of *% Fixation Time to Current* with respect to the *Absolute Duration* differences between the two environments. From this lens, two important findings emerge. First, recall the "just-in-time" effect refers to around half a second of visual fixation on an object/area being required before an individual's end effector arrives to interact there. This "just-in-time" phenomenon can account for a decrease in *% Fixation Time to Current* during Reach and Transport in VR. VR participants move slower, creating more time in each phase of movement. Since the eyes arrive at the same fixed time of about half a second before the hand does, this leads to the reduced % time. Second, the lack of a significant difference in *% Fixation Time to Current* during Release between RW and VR needs to be considered in light of the extremely extended *Absolute Duration* of the Release phase in VR. Participants in both conditions dedicate nearly all their visual attention to the box as it is being released, but this takes much longer in VR.

Eyes fixate significantly more on hand in flight during Transport in VR: One of our key predictions was supported in our results—participants in VR looked more toward their own hand and the object in their hand while they were transporting an object compared to participants in the real world (Figures 5c, d). This occurred at the start of each Transport, which we interpret as participants using eye gaze in VR to ensure they had initiated an object interaction successfully. The first evidence for this is an increase in the *Number of Fixations to the Hand in Flight* (Figure 5c) during the Transport phase (RW = 0.74, *SD* = 0.29; VR = 1.03, *SD* = 0.38; $p = 5.00 \times 10^{-3}$). Since this pattern was not seen in other movement phases (e.g., Reach), this is strong evidence that this difference is not due to differences in eye-tracking calibration or some other hardware or software difference between the real world and VR data sets. The *% Fixation Time to the Hand in Flight* results (Figure 5d) follow a similar pattern. Participants spent much more ($p = 7.56 \times 10^{-5}$) time fixating the Hand in Flight during the Transport phase in VR (23.04%, *SD* = 12.81%) than in the RW (9.43%, *SD* = 5.86%) while there were no statistical differences between condition for the other movement phase (Reach). These percent fixation time results are even larger when considering that the *Absolute Duration* values of the Transport phases in VR were much longer than in the real world. Recall that participants in VR fixate significantly less relative time during Transport on the location they're transporting the object to compared to the RW. This finding gains more clarity knowing that the reason for this is in part that participants are spending more relative time fixating on their own hand, or the object in their hand, during VR, and therefore less fixation time is dedicated to the drop-off location.

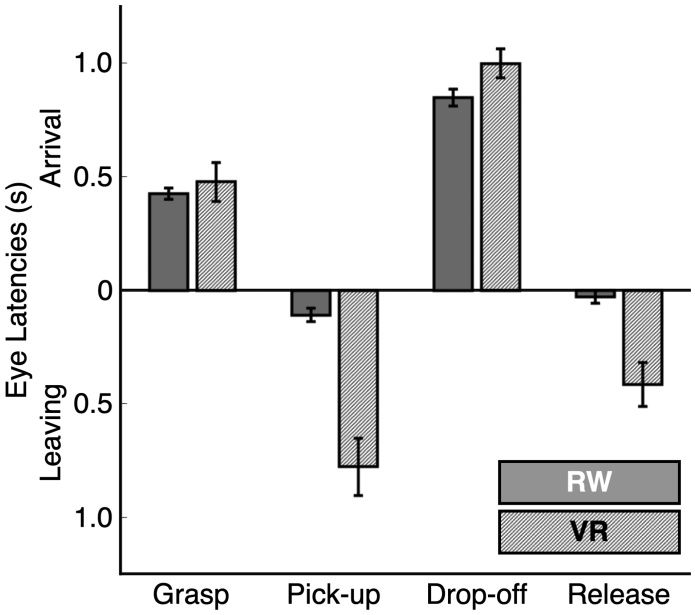

Figure 6. The average of all participants' *EAL Grasp* (s), *ELL PU* (s), *EAL DO* (s), and *ELL Release* (s) of the Pasta Box Task in VR and RW. Error bars represent *SEM*.

Family 3: When do people look?

Visual fixation lingers on objects after pick-up and release in VR: Below we describe the similarities and differences seen between the four eye latency measures in the real world and VR in chronological order (Figure 6). That is, we describe these measures in the order that they would occur as a participant fixates on an object before picking it up (*EAL Grasp*), moving their hand toward the object, picking the object up, shifting their fixation away from the pick-up location (*ELL PU*) and on to the drop-off location (*EAL DO*), and then moving the object toward the drop-off location, dropping off the object, and shifting their fixation away from the object (*ELL Release*).

First, *EAL Grasp* measures the time the eye arrives at the object location prior to the start of an interaction. Here we see the consistent pattern of about 500 ms of advance looking time in both the real world (RW = 0.42s, $SD = 0.10$s) and in VR (VR = 0.48s, $SD = 0.32$s), with no significant differences. Moving chronologically, *ELL Pick-up* was significantly longer in VR than the real world (RW = –0.11 s, $SD = 0.12$ s; VR = –0.78 s, $SD = 0.49$ s; $p = 3.71 \times 10^{-7}$), meaning participants stayed fixating on the object after they had picked it up more than half a second longer in VR than in the real world. This reinforces and refines our main point from the previous section. Not only does lack of haptic information cause participants to visually fixate more on the object during an interaction, but it also concentrates that increased visual fixation at the start of the object interaction.

EAL Drop-off was not significantly different between VR than the real world (RW = 0.85 s, $SD = 0.15$ s; VR = 1.00 s, $SD = 0.23$ s). That is, participants visually fixated for a similar period of time in the real world and VR on the drop-off target before their hand arrived with the object to initiate the drop-off. Finally, *ELL Release* in VR was much longer than in the real world (RW = –0.03 s, $SD = 0.12$ s; VR = –0.42 s, $SD = 0.36$ s; $p = 1.24 \times 10^{-5}$), meaning that while participants in the real world shifted their fixation away from the object almost immediately after their hand finished releasing it, in VR, they stayed fixating on the object for nearly half a second after they had completed the release.

Discussion

This study compares eye–hand coordination during object interactions in real and virtual worlds by re-creating a previously published object interaction task from the real world (Boser et al., 2018; Lavoie et al., 2018; Valevicius et al., 2018, 2019) in controller-mediated VR. Given the lack of haptic feedback, the general assumption is that there would be strong visuomotor differences between controller-mediated VR and the real world. While this is in some ways true (see below), it is also, in important ways, false. To summarize the similarities, we found that participants in VR spent close to the same relative time as their real-world counterparts on the first three phases (Reach, Grasp, Transport) of each object movement. VR participants also fixated objects and targets that were relevant to the immediate task they were doing, or the task they would be doing next, just like real-world participants. Finally, perhaps the most striking similarity was that, like many studies across many different contexts, we found evidence for the "just-in-time" phenomenon of gaze leading the hand. Just like in the previous work including this task in the real world, VR participants using handheld controllers fixated an object/location they were about to interact with a minimum of half a second before their hand arrived.

The consistency of the "just-in-time" hypothesis of eye–hand coordination is remarkable and, importantly, not strictly predicted by the hypothesis that eye–hand coordination in controller-mediated VR would be categorically different than that observed in the real world. Whether it is a person moving a mouse cursor to drag an object on a screen (Bertrand & Chapman, 2023), a monkey catching prey in the wild (Ngo et al., 2022), a person making tea (Land & Hayhoe, 2001), or someone moving a real pasta box (Lavoie et al., 2018) or even a virtual one (current study), successful object interactions seem to require about half a second of visual fixation prior to the initiation of an interaction. Advanced visual fixation helps hand accuracy for several reasons. First, foveating the object allows the combination of multiple signals to locate the object, such as high-resolution visual information, efference copy information from the motor command to the eye muscles, and proprioceptive information from the eye muscles (Bridgeman & Stark, 1991; Poletti, Burr, & Rucci, 2013). Then, visual information of the hand reaching toward an object leads to greater accuracy since humans use vision as a feedback mechanism for the movement system (de Brouwer et al., 2021). An involuntary correction of hand trajectory has been shown in several studies, detailing the tightly coupled nature of vision as the feedback mechanism for reaching movements (de Brouwer et al., 2021; Franklin & Wolpert, 2008; Franklin, Wolpert, & Franklin, 2017; Saunders & Knill, 2003; Wolpert & Flanagan, 2001). Together, this research shows there is an intimate connection between the eyes and hands that can be affected by tiny variations in hand trajectory and gaze location. It seems as though this minimum half a second is necessary to compute grasp dynamics and other required information in advance to the initiation of a successful interaction. Of course, in most movements, there is a trade-off present between speed and accuracy. The faster we move, the less accurate our movements are and the greater our risk of making a mistake. Perhaps this half-second minimum advanced fixation provides the necessary confidence that the task will be completed successfully within a certain reasonable period of time. In the context of the current study, what this highlights is that the mechanisms of visual feedback control to guide a hand toward an object do not appear to be impacted by its "virtuality," suggesting, as others have through neuroimaging (Cavina-Pratesi et al., 2018; Culham et al., 2003), that reach and grasp planning are somewhat dissociable.

As expected, there were differences between the VR and real-world conditions. First, participants in controller-mediated VR took nearly twice as long to complete each trial than their real-world counterparts

with a disproportionate amount of time in the Release phase of each object movement. VR participants spent more relative time visually fixating their hand and the pasta box in their hand during the Transport phase and less relative time fixating on the upcoming drop-off target compared to real-world participants. As well, VR participants spent less relative time fixating on the box during the Reach phase than real-world participants, which can be accounted for by the "just-in-time" phenomenon described above. Finally, we found that in VR, participants held longer fixations on the box while initiating its pick-up before shifting their gaze to the drop-off target compared to real-world participants. And, while dropping off the box in VR, participants fixated on their virtual hand for nearly half a second after the box had been successfully placed on the target, while in the real world, participants ended that fixation almost immediately after the object was successfully placed.

The way VR participants differ from real-world participants in some ways follows the pattern we have previously observed for prosthesis users (Hebert et al., 2019; Marasco et al., 2021). In short, commercial prosthetic limbs do not provide detailed haptic information to the user, except for slight vibrations that may make their way through the end effector to the residual limb, or high-end devices that provide tactile stimulation to the skin of the residual limb or nerve stimulation, neither of which are widespread (Hebert et al., 2019; Marasco et al., 2021). VR participants move slower than their real-world counterparts, just like prosthesis users. They spend more time fixating on their own limb while transporting an object, like prosthesis users. And, when they are starting to move an object (e.g., the end of the "pick-up" interaction) or moving their hand away from an object (e.g., the end of the "drop-off" interaction), VR participants fixate on their hand to ensure the box is doing what they want it to. A lack of haptic feedback means participants don't feel the box as they interact with it. They don't experience the weight of the box as it is lifted up, whether it's gripped tightly or slipping, if it's rotating in their grasp, and if the load is lightened as they go to put it down. As a result, we speculate that both VR and prosthesis users compensate at these key moments by extending visual fixation.

In contrast to what we predicted, participants in VR do not behave like prosthesis users when initiating a grasping movement. Specifically, in our study, we see limited evidence for a prolonged Grasp phase in VR while in prosthesis users, the Grasp phase is significantly lengthened. Again, this highlights an important way that a more simplistic expectation of difference between the real and virtual worlds fails to account for the observed results. Upon reflection, in the VR condition, the vibration of the controller provided an additional haptic feedback cue indicating it was in a position to initiate a grasp and no extra visual attention was required. The feedback vibration likely increased VR participants' confidence in initiating a successful grasp, and the simplicity of pulling the controller trigger meant that the time spent grasping was quite similar to participants in the real world. This stands in striking contrast to the complex grasp mechanisms that many prosthesis users must control (Hebert et al., 2019; Marasco et al., 2021). Unlike in VR, where the vibrating controller guarantees a successful interaction will be initiated, prosthesis users have no such shortcut to confidence and instead must rely on visual feedback through an elongated grasp fixation.

Taken together, these results weave an intricate pattern of how gaze is distributed during object interactions that depends crucially on a participant's confidence that each phase of an interaction will be completed successfully. When one grabs an object, a simple buzz on the hand that perfectly cues interaction success is enough for eye–hand coordination to proceed close to normally, even in the absence of more detailed haptic feedback. But, as one starts to move the object or moves one's hand away from an object after putting it down, participants without haptic feedback appear to rely on extra time spent looking at the site of interaction to attain confidence of its successful completion. This supports the astute observation from Land and Hayhoe (2001) that "vision is a scarce and valuable resource, and it is disengaged from a particular aspect of an action as soon as another sense is available to take over." In a typical real-world interaction, that other sense is haptic and proprioceptive sensory information from the hand. In situations with reduced haptic feedback like controller-mediated VR or using a prosthesis, with no other sense "available to take over," vision is required to linger longer at the site of an interaction.

Theoretically, this also aligns with the attentional landscapes theory of visual attention (Baldauf & Deubel, 2010), which explains the deployment of overt and covert attention across time and space. Previously, we explained eye–hand coordination in the real world using the attentional landscapes theory (Lavoie et al., 2018). The shift of visual fixation from the current, overtly attended action site to the future covertly attended action site occurs fluidly and quite close to the onset and termination of object movement, as observed by *ELL Pick-up* and *ELL Release* being quite close to zero. Here, we propose that in controller-mediated VR, confidence is compromised at the current site of action and necessitates maintaining overt attention. This delay stalls the increase of covert attention at the next site of action, leading to an attentional landscape that is unable to shift fluidly, resulting in the reported *ELL Pick-up* being nearly a full second longer and *ELL Release* being nearly half a second longer than in the real world. The halting of the fluid flow of the attentional landscape likely contributes to the slowdown in VR compared to

the real world during object interactions. It's not in the participant's best interests to get the box to the drop-off target too quickly, as the eyes won't have visually fixated the drop-off target for long enough to compute the appropriate drop-off motor dynamics. Participants, therefore, slow down their movement, leaving enough time for successful drop-off computation.

While a lack of haptic feedback is likely the dominant explanation for the differences we observed, there are a number of other possible factors that could be contributing to the reported effects. First, VR participants had to wear a headset, while the real-world participants only had to wear a head-mounted eye tracker. Wearing a heavier headpiece may have caused VR users to move slower. The effects of a head that cannot orient as quickly could also explain some of our results. Second, as mentioned earlier in this article, it's been shown that VR objects are treated differently than real-world objects, which could have caused some of the differences we've found here. While we feel the consistency of the 500-ms advance look time when reaching runs counter to this idea, the overall slowness of movements in VR may in part be due to acting on virtual object proxies. Finally, in this experiment, VR users actually received a small vibration on their hand when they were able to initiate a grasp, meaning there was some haptic feedback. This highlights the need to truly explore a situation with full and fully deprived haptic feedback to more conclusively isolate our results to that modality. For the above reasons and more, we plan to pursue further investigation that tests the precise changes that occur in visuomotor behavior in VR when full haptic feedback is provided to users' hands during object interactions, compared to when no or minimal haptic feedback is provided.

Conclusions

In some ways, these results show that participants in controller-mediated VR doing the same task behave quite similarly to those in the real world—they look at the same objects and locations in the same order and for about the same amount of relative time. Most strikingly, their eye gaze seems to follow the "just-in-time" phenomenon seen in myriad contexts—the eyes arrive at the site of an interaction at least 500 ms before the interaction begins. But, despite these similarities, there are also important differences. Participants take nearly twice as long to complete the set of movements in controller-mediated VR than in the real world, especially when releasing an object after moving it. We speculate that this difference is mainly driven by a lack of haptic feedback. Deprived of haptics, VR participants become more reliant on vision to confirm that an interaction has been completed successfully. Thus, their eyes look more toward their hand as it moves an object and lingers at the site of interaction longer, putatively necessitating slower movements.

Taken together, this study shows that, although VR is a useful tool, its use for skill training is unlikely to be effective for visuomotor behaviors. Until accurate haptic feedback in VR is accomplished, we expect to see compensatory strategies, particularly invocation of greater visual attention during specific portions of object interactions, to make up for the lack of the rich, touch information that our hands so importantly provide. If the intention is to increase the performance of users' dexterous eye–hand behavior in the real world, the lack of haptic feedback from most VR technology could do more harm than good as users learn a nonoptimal compensatory strategy. This last point highlights the acute need to develop better and more standardized assessments of skilled motor performance in virtual and augmented realities. As new VR technologies are developed, including hand-tracking and haptic feedback devices, we suggest that eye–hand coordination metrics during object interactions—like those reported in this study—should be used as an important benchmark to see how close users of new devices are with respect to real-world visuomotor behavior.

Keywords: virtual reality, eye–hand coordination, eye tracking, visual attention, object interactions

Acknowledgments

The authors thank the efforts of Riley Dawson and Quinn Boser in the design and development of the data analysis tool, GaMA. The authors thank these people and organizations for their ongoing support, without which this research would not have been possible.

E.L. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) with a Canada Graduate Scholarship–Doctoral (CGSD3-519162-2018) and is currently supported by the TD Bank Financial Group Grant for Health Sciences Interdisciplinary Research Fund Award. C.C. receives funding through an NSERC Discovery Grant (RGPIN-2020-05396), the Canadian Foundation for Innovation John R. Evans Leaders Fund, and a Canadian Institute for Advanced Research Catalyst Grant.

Commercial relationships: none.
Corresponding author: Ewen Lavoie.
Email: elavoie@ualberta.ca.
Address: Faculty of Kinesiology, Sport, and Recreation, Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada.

References

Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research, 50*(11), 999–1013.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80, https://doi.org/10.1162/jocn.1995.7.1.66.

Bertrand, J. K., & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

Boser, Q. A., Valevicius, A. M., Lavoie, E. B., Chapman, C. S., Pilarski, P. M., Hebert, J. S., . . . Vette, A. H. (2018). Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. *Journal of Biomechanics, 72*, 228–234, https://doi.org/10.1016/j.jbiomech.2018.02.028.

Bridgeman, B., & Stark, L. (1991). Ocular proprioception and efference copy in registering visual direction. *Vision Research, 31*(11), 1903–1913, https://doi.org/10.1016/0042-6989(91)90185-8.

Cavina-Pratesi, C., Connolly, J. D., Monaco, S., Figley, T. D., Milner, A. D., Schenk, T., . . . Culham, J. C. (2018). Human neuroimaging reveals the subcomponents of grasping, reaching and pointing actions. *Cortex, 98*, 128–148.

Cisek, P. (2022). Evolution of behavioural control from chordates to primates. *Philosophical Transactions of the Royal Society B, 377*(1844), 20200522.

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., . . . Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review, 23*(2), 640–647.

Culham, J. C., Danckert, S. L., DeSouza, J. F. X., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research, 153*(2), 180–189.

de Brouwer, A. J., Flanagan, J. R., & Spering, M. (2021). Functional use of eye movements for an acting system. *Trends in Cognitive Sciences, 25*(3), 252–263.

Desmurget, M. (1998). From eye to hand: Planning goal-directed movements. *Neuroscience & Biobehavioral Reviews, 22*(6), 761–788, https://doi.org/10.1016/s0149-7634(98)00004-9.

Franklin, D. W., & Wolpert, D. M. (2008). Specificity of reflex adaptation for task-relevant variability. *Journal of Neuroscience, 28*(52), 14165–14175, https://doi.org/10.1523/jneurosci.4406-08.2008.

Franklin, S., Wolpert, D. M., & Franklin, D. W. (2017). Rapid visuomotor feedback gains are tuned to the task dynamics. *Journal of Neurophysiology, 118*(5), 2711–2726.

Goodale, M. A., Jakobson, L. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia, 32*(10), 1159–1178.

Harris, D. J., Buckingham, G., Wilson, M. R., & Vine, S. J. (2019). Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. *Experimental Brain Research, 237*(11), 2761–2766.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188–194, https://doi.org/10.1016/j.tics.2005.02.009.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1), 6–6.

Hebert, J. S., Boser, Q. A., Valevicius, A. M., Tanikawa, H., Lavoie, E. B., Vette, A. H., . . . Chapman, C. S. (2019). Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA Network Open, 2*(9), e1911197, https://doi.org/10.1001/jamanetworkopen.2019.11197.

Hebert, J. S., & Shehata, A. W. (2022). The effect of sensory feedback on the temporal allocation of gaze using a sensorized myoelectric prosthesis. *Proceedings of the Myoelectric Control Symposium*, https://www.unb.ca/ibme/_assets/documents/mec22proceedings2.pdf.

Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific Reports, 6*, 24528.

Johansson, R. S., Westling, G., Bäckström, A., & Randall Flanagan, J. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience, 21*(17), 6917–6932, https://doi.org/10.1523/jneurosci.21-17-06917.2001.

Joyner, J. S., Vaughn-Cooke, M., & Benz, H. L. (2021). Comparison of dexterous task performance in virtual reality and real-world environments. *Frontiers in Virtual Reality, 2*, https://doi.org/10.3389/frvir.2021.599274.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience, 26*(1), 51–62, https://doi.org/10.1017/s0952523808080899.

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., . . . Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision, 18*(6), 18, https://doi.org/10.1167/18.6.18.

Lavoie, E., & Chapman, C. S. (2021). What's limbs got to do with it? Real-world movement correlates with feelings of ownership over virtual arms during object interactions in virtual reality. *Neuroscience of Consciousness,* https://doi.org/10.1093/nc/niaa027.

Lerner, D., Mohr, S., Schild, J., Göring, M., & Luiz, T. (2020). An immersive multi-user virtual reality for emergency simulation training: Usability study. *JMIR Serious Games, 8*(3), e18822.

Levac, D. E., Huber, M. E., & Sternad, D. (2019). Learning and transfer of complex motor skills in virtual reality: A perspective review. *Journal of Neuroengineering and Rehabilitation, 16*(1), 121.

Mao, R. Q., Lan, L., Kay, J., Lohre, R., Ayeni, O. R., Goel, D. P., . . . de Sa, D. (2021). Immersive virtual reality for surgical training: A systematic review. *The Journal of Surgical Research, 268*, 40–58.

Marasco, P. D., Hebert, J. S., Sensinger, J. W., Beckler, D. T., Thumser, Z. C., Shehata, A. W., . . . Wilson, K. R. (2021). Neurorobotic fusion of prosthetic touch, kinesthesia, and movement in bionic upper limbs promotes intrinsic brain behaviors. *Science Robotics, 6*(58), eabf3368.

Marini, F., Breeding, K. A., & Snow, J. C. (2019). Distinct visuo-motor brain dynamics for real-world objects versus planar images. *NeuroImage, 195*, 232–242.

Milner, D., & Goodale, M. (2006). *The visual brain in action* (Vol. 27). OUP Oxford.

Neggers, S. F. W., & Bekkering, H. (2000). Ocular gaze is anchored to the target of an ongoing pointing movement. *Journal of Neurophysiology, 83*(2), 639–651, https://doi.org/10.1152/jn.2000.83.2.639.

Ngo, V., Gorman, J. C., De la Fuente, M. F., Souto, A., Schiel, N., & Miller, C. T. (2022). Active vision during prey capture in wild marmoset monkeys. *Current Biology: CB, 32*(15), 3423–3428.e3.

Oagaz, H., Schoun, B., & Choi, M.-H. (2022). Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on Visualization and Computer Graphics, 28*(12), 4332–4343.

Poletti, M., Burr, D. C., & Rucci, M. (2013). Optimal multimodal integration in spatial localization. *The Journal of Neuroscience, 33*(35), 14259–14268.

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal, 6*(3), 181–185, https://doi.org/10.7861/fhj.2019-0036.

Saunders, J. A., & Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental Brain Research, 152*(3), 341–352.

Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Sciences, 25*(6), 506–519.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), 5–5.

Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Chapman, C. S., Pilarski, P. M., Hebert, J. S., . . . Vette, A. H. (2019). Characterization of normative angular joint kinematics during two functional upper limb tasks. *Gait & Posture, 69*, 176–186, https://doi.org/10.1016/j.gaitpost.2019.01.037.

Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Murgatroyd, G. S., Pilarski, P. M., Chapman, C. S., . . . Hebert, J. S. (2018). Characterization of normative hand movements during two functional upper limb tasks. *PLoS ONE, 13*(6), e0199549, https://doi.org/10.1371/journal.pone.0199549.

Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H., & Hebert, J. S. (2019). Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. *PLoS One, 14*(12), e0219333.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology: CB, 11*(18), R729–R732.