

Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks

Ewen B. Lavoie

Faculty of Kinesiology, Sport, and Recreation,
University of Alberta, Edmonton, Alberta, Canada



Aida M. Valevicius

Department of Biomedical Engineering,
University of Alberta, Edmonton, Alberta, Canada



Quinn A. Boser

Department of Biomedical Engineering,
University of Alberta, Edmonton, Alberta, Canada



Ognjen Kovic

Division of Physical Medicine and Rehabilitation,
Department of Medicine, University of Alberta,
Edmonton, Alberta, Canada



Albert H. Vette

Department of Mechanical Engineering,
University of Alberta, Edmonton, Alberta, Canada
Department of Biomedical Engineering,
University of Alberta, Edmonton, Alberta, Canada
Glenrose Rehabilitation Hospital,
Alberta Health Services, Edmonton, Alberta, Canada
Neuroscience and Mental Health Institute,
University of Alberta, Edmonton, Alberta, Canada



Patrick M. Pilarski

Division of Physical Medicine and Rehabilitation,
Department of Medicine, University of Alberta,
Edmonton, Alberta, Canada



Jacqueline S. Hebert

Division of Physical Medicine and Rehabilitation,
Department of Medicine, University of Alberta,
Edmonton, Alberta, Canada
Department of Biomedical Engineering,
University of Alberta, Edmonton, Alberta, Canada
Glenrose Rehabilitation Hospital,
Alberta Health Services, Edmonton, Alberta, Canada
Neuroscience and Mental Health Institute,
University of Alberta, Edmonton, Alberta, Canada



Craig S. Chapman

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



Citation: Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & 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, 1–20, <https://doi.org/10.1167/18.6.18>.



This study explores the role that vision plays in sequential object interactions. We used a head-mounted eye tracker and upper-limb motion capture to quantify visual behavior while participants performed two standardized functional tasks. By simultaneously recording eye and motion tracking, we precisely segmented participants' visual data using the movement data, yielding a consistent and highly functionally resolved data set of real-world object-interaction tasks. Our results show that participants spend nearly the full duration of a trial fixating on objects relevant to the task, little time fixating on their own hand when reaching toward an object, and slightly more time—although still very little—fixating on the object in their hand when transporting it. A consistent spatial and temporal pattern of fixations was found across participants. In brief, participants fixate an object to be picked up at least half a second before their hand arrives at the object and stay fixated on the object until they begin to transport it, at which point they shift their fixation directly to the drop-off location of the object, where they stay fixated until the object is successfully released. This pattern provides additional evidence of a common system for the integration of vision and object interaction in humans, and is consistent with theoretical frameworks hypothesizing the distribution of attention to future action targets as part of eye and hand-movement preparation. Our results thus aid the understanding of visual attention allocation during planning of object interactions both inside and outside the field of view.

Introduction

Humans are extremely effective at interacting with objects to accomplish daily tasks, and these countless interactions happen so seamlessly we are not aware of the complex integration of sensory modalities that must be occurring. In particular, the vestibular system provides information about balance and spatial orientation, proprioception helps us move our limbs effectively, and haptic feedback from the hands informs us when we manipulate an object. Vision plays arguably the most dominant role in efficient object interaction, with clear evidence that the dorsal visual stream plays a critical role in generating the motor plans driving hand and arm movements during visually guided object interactions (Desmurget, Pélisson, Rossetti, & Prablanc, 1998; Milner & Goodale, 2006). Typically, the eyes fixate on the target of action during a pointing movement (Neggers & Bekkering, 2000) and on key areas like obstacles or objects during object-manipulation tasks, but rarely on a participant's own hand (Johansson, Westling, Bäckström, & Flanagan, 2001). Quantifying the allocation of visual attention can provide important insights into movement plan-

ning, and is most commonly achieved by using metrics derived from eye-movement behavior as a predictor of visual attention.

Eye movements have long been used as a probe of how people gain information from the world (Kowler, 2011). Early studies found that eye-movement patterns during free viewing of pictures can differ substantially between individuals, but when participants are given a specific goal to accomplish, their eye-movement patterns become more similar (Buswell, 1935; Yarbus, 1967) and shift based on the specific task (e.g., estimate the age versus remember the clothing of people in a scene). However, this top-down control of eye movements for goal-directed tasks can be affected by the testing scenario. The majority of studies of eye behavior use static or remote monitoring with restrictive lab-based tasks (Richardson & Spivey, 2008). Inhibiting movements of the head and body during experimentation has been shown to affect eye behavior; for example, the velocity of saccades is higher if the head is unable to turn freely toward the targeted future fixation point (Freedman, 2008). It has also been shown that visuomotor behavior differs when participants are asked to perform a motor action (tapping the finger) in addition to visually fixating on objects compared to fixating on objects without an accompanying motor action (Epelboim et al., 1997). This reinforces the limitations of lab-based tasks (e.g., looking at or touching shapes on a screen) in representing the demands of the eye-movement system in the real world (Kingstone, Smilek, & Eastwood, 2008).

In contrast, head-mounted eye trackers have helped researchers design experiments that examine eye-movement behavior during everyday tasks in natural settings by allowing participants to move their heads, bodies, and eyes more normally (Holmqvist et al., 2011). This has amounted to an increase in literature about human behavior in more natural, everyday tasks—like driving (Land & Lee, 1994), handwashing (Pelz & Canosa, 2001), and preparing a sandwich (Hayhoe, 2000; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003) or cup of tea (Land, Mennie, & Rusted, 1999)—leading to important generalizations about eye behavior during complex tasks (Land & Hayhoe, 2001; Hayhoe et al., 2003; Hayhoe & Ballard, 2005; Land, 2009; Tatler, Hayhoe, Land, & Ballard, 2011). For example, we know that the eyes strongly precede and predict almost every action, suggesting that the visuomotor system solves object-interaction problems in real time (Land & Hayhoe, 2001; Land, 2009; Tatler et al., 2011). In addition, monitoring eye behavior during complex tasks allows assessment of the impact of internal reward systems (Hayhoe & Ballard, 2005; Tatler et al., 2011) and implicit memory structures (Chun & Nakayama, 2000; Hayhoe et al., 2003) on

optimal planning and coordination of eye and hand movements in the real world. However, providing freedom to participants to accomplish the goals of an open-ended complex task can result in higher variability in performance, as each participant may carry out the task in a different order (Land et al., 1999). This allows some generalizability of eye-behavior, but specific temporal dynamics between the eyes, head, and body are not able to be systematically observed. For example, Land and Hayhoe (2001) showed that participants make a saccade away from an object being grasped to its future drop-off location but were not able to identify the exact timing of this saccade relative to when the object began moving. Thus, although important and informative, these studies provide only a coarse level of functional resolution of eye-movement behavior during object-interaction tasks.

As the main contribution of this work, we leverage advancements not only in head-mounted eye tracking but also in synchronized and simultaneously recorded motion tracking to offer a higher-functional-resolution study of eye-movement behavior during object interactions representative of real-world tasks. In particular, accurate information about the locations of objects and about body movements allows a determination of exactly where a person is looking at each point leading up to, and following, an object interaction. Thus, objective segmentation of eye-movement data into the phases of each body movement can occur. We hypothesized that providing participants goal-oriented tasks to accomplish in a specific order would enable us to observe consistent eye-movement behavior across participants and across multiple repetitions of the task. In turn, unlike previous work, this consistency and segmentation of movements would allow us to be more specific in our generalizations across task behavior and to offer a more precise description of eye behavior during object interaction.

The goals of our study were therefore to examine eye-movement behavior of participants performing multiple repetitions of standardized simulated real-world tasks. Specifically, we aimed to do the following:

- Use two standardized goal-based tasks involving defined object interactions that are representative of real-world tasks.
- Use the least restraining eye-tracking technology available, to allow participants to perform these tasks in the most natural way possible.
- Segment eye data objectively using motion-tracking information to standardize the eye-behavior description and determine temporal dynamics.
- Derive general principles governing eye-movement behavior during object-related actions.

Thus, we recorded eye and movement behavior during two tasks mimicking real-world demands,

establishing a normative data set for functional eye-gaze behavior. The first task emulates moving a box of pasta from a countertop into a cupboard, and the second emulates moving filled cups across a countertop. Motion tracking allowed us to objectively segment object interactions into *Reach*, *Grasp*, *Transport*, and *Release* phases and, thus, to compare how long people spend in each of these phases and where they are looking during each phase. Moreover, since this study is primarily interested in quantifying eye behavior during object interaction, we further defined two key events (based on the tracked movement of the object): *Pick-up*, referring to the transition from a Grasp to a Transport as the object starts being moved, and *Drop-off*, referring to the transition from Transport to Release as the object stops moving. Using these movement-defined events, we aimed to uncover the temporal dynamics between the location of visual fixation and the location of the hand and objects. Specifically, we calculated the difference in time between an object being visually fixated during Pick-up and the hand beginning a manipulation of the object, as well as the difference between the Drop-off location being fixated and the hand releasing the object at that location. Finally, we also calculated when the transition of fixation from Pick-up to Drop-off occurred.

In general, we hypothesized we would find the same high-level pattern of results as has been reported previously (Land & Hayhoe, 2001; Hayhoe & Ballard, 2005; Land, 2009; Tatler et al., 2011): Participants would almost exclusively fixate task-relevant objects, rarely fixate their own hand, and have their eyes lead the hand by about half a second. Additionally, with the precise segmentation afforded by our integration of motion capture and the comparatively large number of trial repetitions, we expected to demonstrate the consistency of these high-level properties across different movement types.

In brief, this study showed that across two tasks participants spent similar amounts of time completing specific phases of movements (e.g., the time spent transporting was similar across both tasks). For eye movements, the vast majority of fixations were to objects and areas relevant to the completion of the task, with participants beginning the fixation of an object to be interacted with more than half a second before the hand arrived. This pattern of the eyes leading the hand meant that participants spent very little time fixating on their own hand, except at the beginning moments of transporting an object. Finally, physical task and biomechanical constraints led to participants fixating less on objects that were outside their field of view, having their eyes arrive later to objects when in certain anatomical configurations (e.g.,

arm across body), and looking earlier at objects requiring a grasp interaction.

Methods

Participants

A group of 24 adults, with no upper-body pathology or history of neurological or musculoskeletal injuries within the past 2 years, provided written informed consent to participate in our study. Of these, four data sets were dropped due to apparatus and/or software issues. The remaining 20 participants (11 male, nine female) had an average age of 25.8 ± 7.2 years and an average height of 173.8 ± 8.3 cm, and were made up of 18 self-reported preferred right-hand users and two self-reported preferred left-hand users. Eighteen participants had normal or corrected-to-normal vision, while two participants were tested without corrected vision, as they removed their glasses to don the eye tracker. These two 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.

Apparatus

Participants were fitted with a head-mounted, binocular eye tracker (Dikablis Professional 2.0, Ergoneers GmbH, Manching, Germany), which can be seen in Figures 1 and 2. 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 placed on the participant 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 on the pasta box, cups, and other task-relevant parts of the apparatus (see Figures 1 and 2 and Supplementary Material S1 and S2). The purpose of the motion-capture markers was to track the hand and objects to allow consistent segmentation of the data across participants, and to reconstruct each participant's data in the 3-D virtual environment of the motion-tracking coordinate frame, including a gaze vector showing their visual fixation behavior.

Procedure

Experimental setup

The Dikablis headset recorded pupil movements in infrared at 60 Hz, and was equipped with a forward-facing, high-definition scene camera that recorded the participant's first-person view. The cameras were first optimally positioned and then calibrated (using the Dikablis DLab software), enabling experimenters to position the cameras for the best data collection specific to the participant and task. If any of the cameras were moved after calibration, or if the headset shifted, the calibration process was repeated. Two gaze and motion calibration trials were carried out immediately before data collection to combine the eye and motion data post hoc. For these, participants were instructed to maintain visual fixation on a motion-capture marker attached to the tip of a calibration wand as the experimenter moved the marker through the task space for approximately 90 s.

Functional tasks

For full task details, including apparatus drawings and participant instructions, please refer to the Supplementary Material S1 and S2: Pasta Box Transfer Task and Cups Transfer Task descriptions. The order of the two standardized functional tasks was randomized. Each task was completed as many times as necessary to obtain 20 trials without errors. Both tasks were performed by the right hand only, and participants were asked to keep their left hand in a relaxed position. The tasks were performed under three conditions (eye tracking only, motion tracking only, and both eye and motion tracking), for a total of 60 trials per participant per task. Since this article relies on the combined eye- and motion-tracking data, we report results only from the 20 trials for each task when participants were wearing both the motion-tracking markers and the eye tracker.

Pasta box transfer task: The Pasta box transfer task (Pasta task) consisted of three object movements, all starting and ending with the hand on a standard *Home* position and eyes fixated on a motion-capture marker in the center of the shelf frame (*Neutral* position). First, participants moved a pasta box from the *Start/End Target* on a side cart at the right side of the body onto a *Mid Shelf Target* in front of them (Movement 1; Figure 1a). Then they moved the box from the *Mid Shelf Target* to the *High Shelf Target* by crossing the body's midline (Movement 2; Figure 1b). Finally, they picked up the pasta box from the *High Shelf Target* and placed it back at its initial position on the *Start/End Target* on the side cart (Movement 3; Figure 1c). After each drop-off of the pasta box, participants touched the *Home* position before initiating the next movement of the

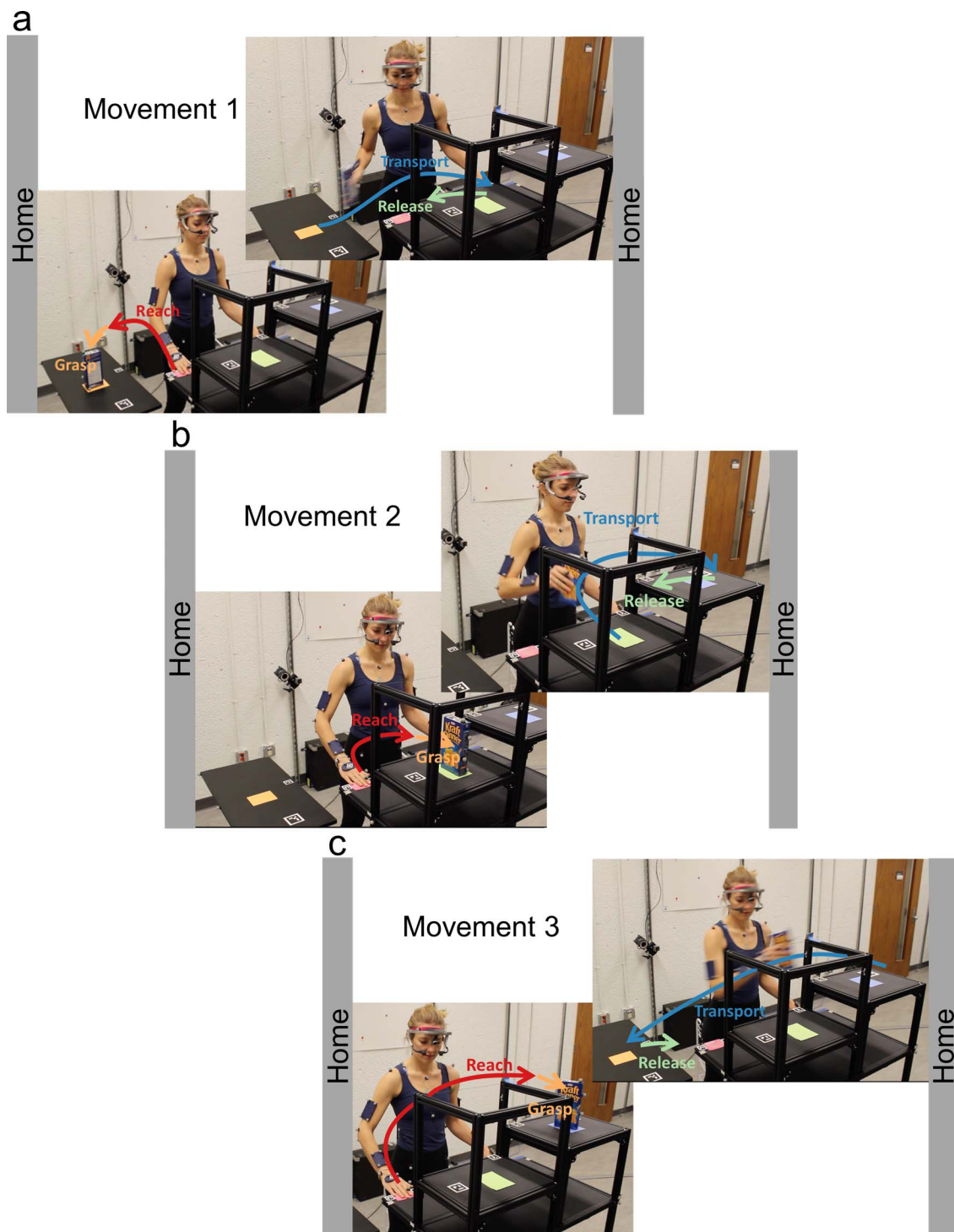


Figure 1. The Pasta box transfer task includes Reach, Grasp, Transport, and Release of a pasta box at three target locations. (a) Movement 1: Grasp from side cart (Start/End Target) and Release on Mid Shelf Target. (b) Movement 2: Grasp from Mid Shelf Target and Release on High Shelf Target. (c) Movement 3: Grasp on High Shelf Target and Release on Start/End Target.

pasta box. Specific dimensions of the Pasta task setup can be found in the Supplementary Material S1.

Participants were instructed to perform the task at a comfortable but efficient pace (for full instructions, please refer to the Supplementary Material S1 (Pasta Task Description)). At each placement endpoint, there

was a colored target indicating where the box should be placed, and participants were told to place the box on the short edge within the boundaries of each placement target. Additionally, they were told to avoid dropping the box, contacting the apparatus, hesitating, or making undesired movements (like scratching their

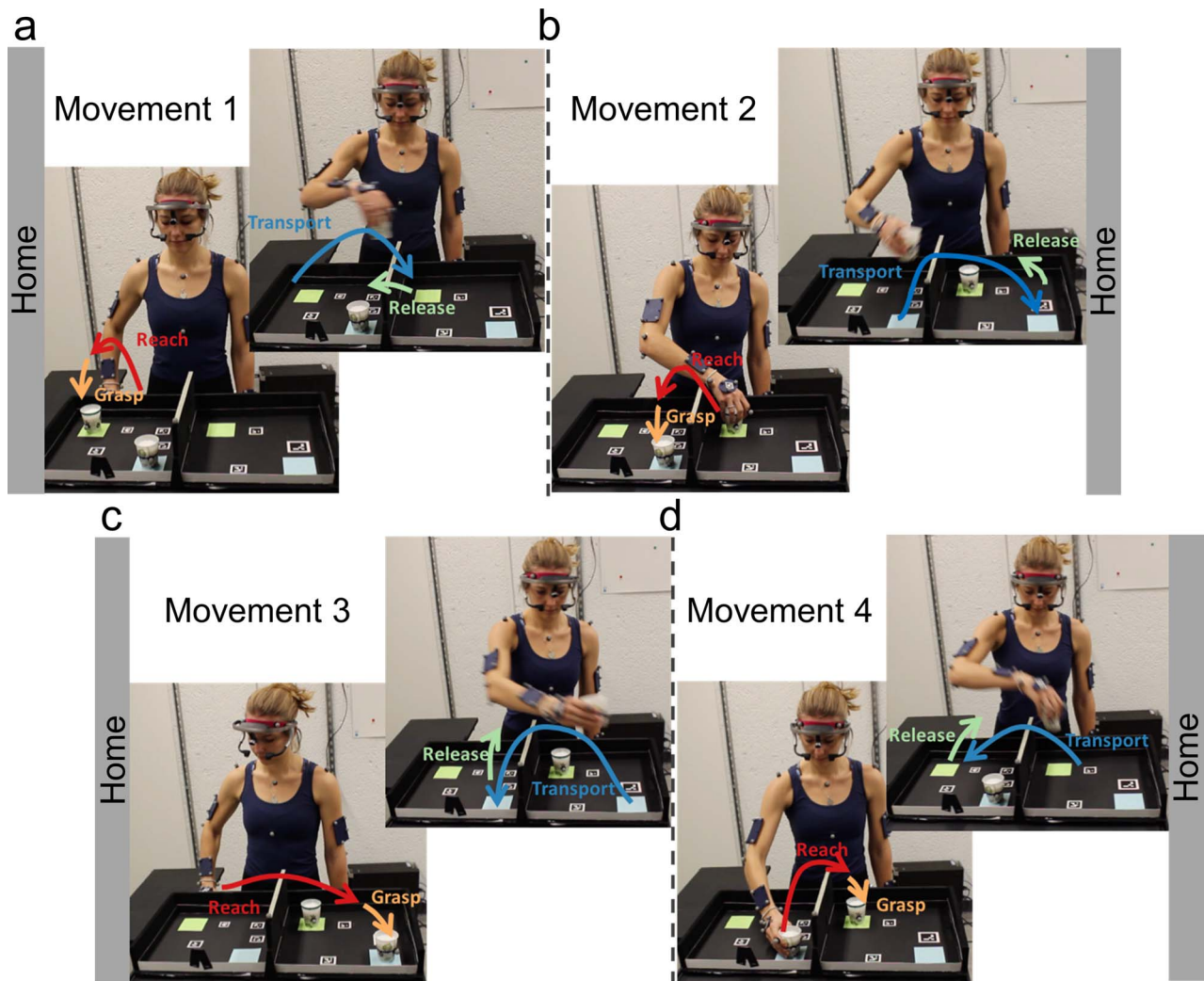


Figure 2. The cup transfer task includes Reach, Grasp, Transport, and Release of two cups at four target locations. (a) Movement 1: Grasp of the green cup with a top grasp at Near Target 1 and Release at Near Target 2. (b) Movement 2: Grasp of the blue cup with a side grasp at Far Target 1 and Release at Far Target 2. (c) Movement 3: Grasp of the blue cup with a side grasp at Far Target 2 and Release at Far Target 1. (d) Movement 4: Grasp of the green cup with a top grasp at Near Target 2 and Release at Near Target 1.

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. For example, if a participant violated one of these rules in three separate trials, 23 trials were collected. On average, each participant performed 22.7 trials of the Pasta task.

Cup transfer task: The Cup transfer task (Cups task) consisted of moving two cups filled with beads (simulating being filled with liquid) over a partition and back again for a total of four object movements. The cups were deformable and would spill beads if grabbed too hard. Like in the Pasta task, participants started each trial with their hand on the Home position and their eyes fixated on a centered motion-capture marker (*Neutral* position). The first and second movements moved the two cups from right to left over the partition. First, a cup with a green rim (the green cup)

was moved from a near right location (closer to the body of the participant; (*Near Target 1*) over the partition to a near left position (*Near Target 2*) with a top grasp (Movement 1; see Figure 2a). Second, a cup with a blue ring around its center (the blue cup) was moved from a far right position (*Far Target 1*) to a far left position (*Far Target 2*) with a side grasp (Movement 2; see Figure 2b). At the end of Movement 2, participants returned their hand to the Home position and proceeded to transport the cups back to their initial positions by reversing the order, therefore moving the blue cup first (from *Far Target 2* to *Far Target 1*; Movement 3; Figure 2c) and the green cup second (from *Near Target 2* to *Near Target 1*; Movement 4; Figure 2d), returning their hand to the Home position after Movement 4. Specific dimensions of the Cups task setup can be found in the Supplementary Material S2.

Along with the grasp instructions that have been outlined, participants were asked to perform the Cups task with similar rules to the Pasta task, but also to avoid deforming the cups or spilling any beads (for full instructions, please refer to the Supplementary Material S2 (Cups Task Description)). If a rule was violated, participants were instructed to complete the trial to the best of their ability, and an extra trial was added at the end of that group of trials. On average, each participant performed 23.4 trials of the Cups task.

Data processing

To synchronize the eye- and motion-tracking data collection, custom software was created to trigger the start and end of the recordings simultaneously, and postprocessing was performed to align the two data streams. Any trial with a difference in the durations of the eye- and motion-tracking recordings greater than 0.400 s was discarded. The average difference for correction was 0.124 s. Custom MATLAB scripts were written to create a regression function using the x - and y -coordinates of each eye in the video frame of the eye tracker and the motion-capture markers on the head and tracked object from the gaze and motion calibration trials. This regression function was then applied to the synchronized eye and motion task data, yielding 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. In some cases the gaze vector was not able to be reconstructed, due to missing or poor data in the gaze calibration file, resulting in a total of 24 trials (out of 400) being discarded from the Cups task, including one full participant, and 89 trials (out of 400) discarded from the Pasta task, including three full participants.

Data segmentation

To identify each object movement and to further segment each Movement into its Reach, Grasp, Transport, and Release phases, we used the motion capture data to conduct the following steps. First, the velocities of the hand and objects were calculated (see Figures 3 and 4 for examples of hand and object velocity traces), as well as the grip aperture of the hand (distance between the thumb and index-finger markers), for all participant trials. For the hand, velocity was calculated for the average position of the three markers attached to the rigid hand plate. For the pasta box, velocity was calculated for the center of a rectangular prism that accurately matched the size and position of the real box, and followed the translation and rotation of the four markers attached to the box. Finally, for the green and blue cups, velocity was calculated for the single marker attached to the back (e.g., away from the participant) of

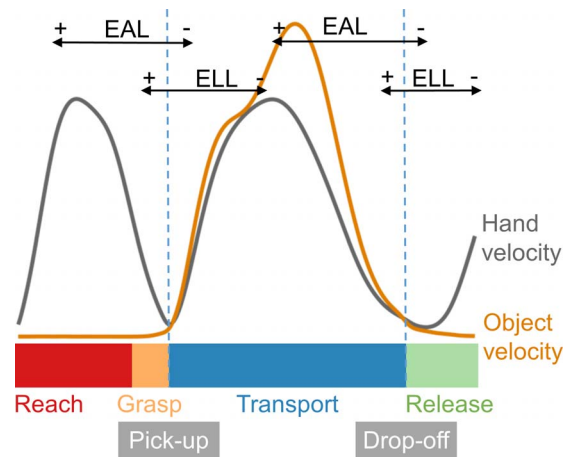


Figure 3. The segmentation of an object movement into its Reach, Grasp, Transport, and Release phases was determined by the velocity of the object (orange trace), the velocity of the hand (gray trace), and grip aperture. Also shown are the approximate temporal locations defined by the terms Pick-up and Drop-off, and the eye-arrival and eye-leaving latency measures associated with each.

each of these objects. While this is not analyzed in this article, it is interesting to note that the objects have higher velocities than the hand, even though they move as a unit. This is due to the arm and wrist rotations that occur during Transport, combined with the more distal position of the object. The start of the Reach phase was defined as the first hand movement (determined by hand velocity) before the subsequent Grasp, ending when the Grasp started. The Grasp phase began when the hand fell within a threshold distance of the object and ended at the onset of Transport. The distance threshold was defined by the distance between the hand and the object at the point of peak grip aperture prior to Transport, averaged across all participants. The onset and offset of object movement (determined by object velocity) defined the start and end of the Transport phase (Figure 3). Finally, the Release phase began when Transport ended and continued until the hand left a threshold distance from the object. This distance threshold was defined by the distance between the hand and object at the point of peak release aperture after Transport, averaged across all participants. The grasp and release distance values were set separately per Pick-up/Drop-off location and are listed in the Supplementary Figure S1. We elected to use these fixed distances across participants instead of individual grip apertures to allow more consistency in dealing with participants whose grip-aperture data were less reliable.

Areas of interest

Because we were interested in overt fixations to areas relevant to object interactions, and since previous

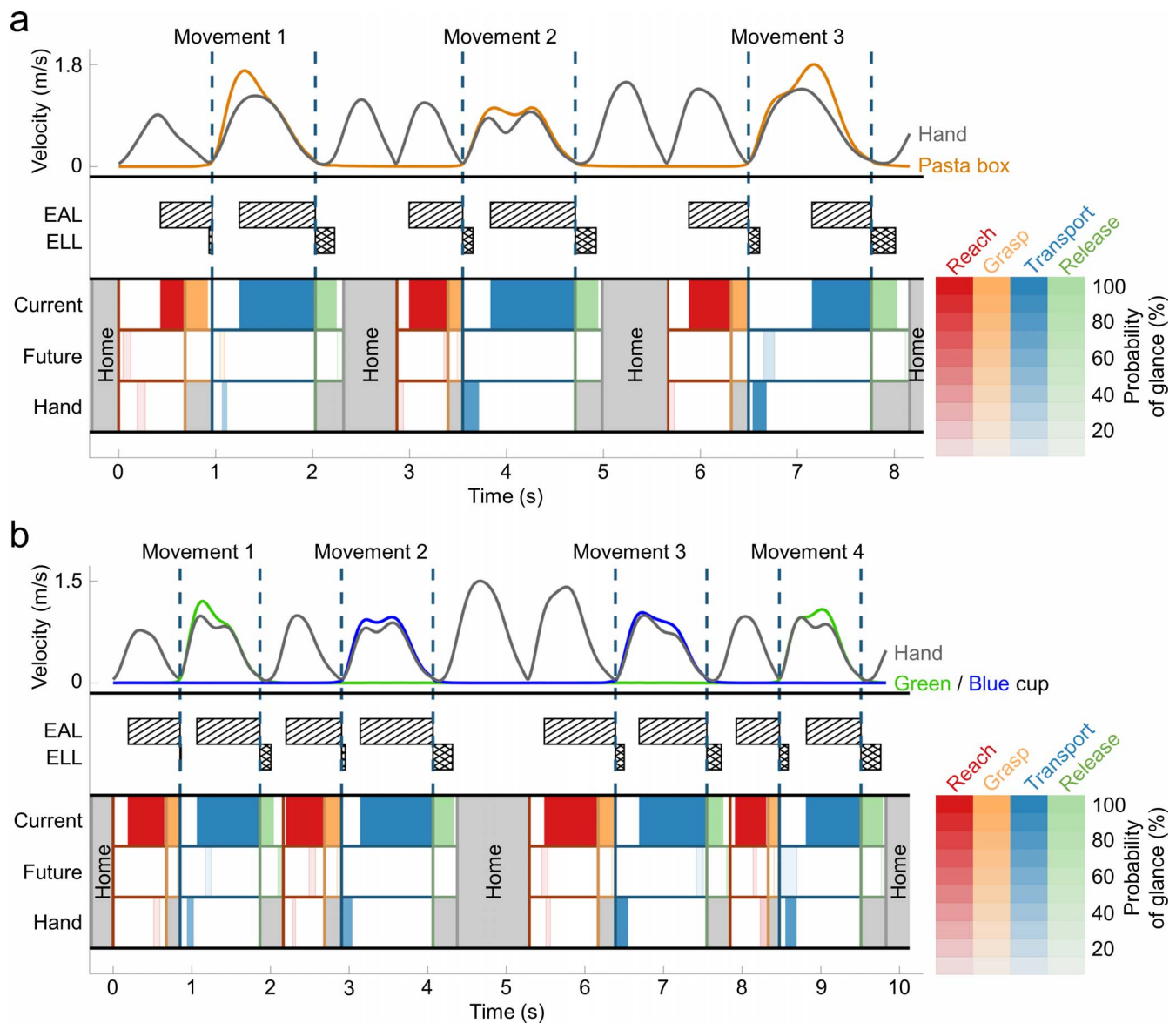


Figure 4. Average timeline of hand and object velocities (top plots), eye-arrival and -leaving latencies (middle plots), and fixations to areas of interest (bottom plots) for (a) the Pasta task and (b) the Cups task. (a) The Pasta task was divided into three movements based primarily on hand (gray) and pasta-box (orange) velocities (top plot). The eye arrived at the interaction location (EAL, slanted fill) well before an object was picked up or dropped off, and usually left just after Pick-up or Drop-off (ELL, hatched fill; middle plot). Movements were subdivided into Reach (red outlines), Grasp (orange outlines), Transport (blue outlines), and Release (green outlines) phases (bottom plot, returns to Home shown in gray), and fixations (color within outlines) were recorded toward the Current interaction location (top bars), the next interaction location (Future, middle bars), or exclusively the hand or object in hand in flight (Hand, bottom bars). Percent fixation time is the duration of fixation relative to the duration of the phase, and the number of fixations is denoted by probability (legend at bottom right). (b) The Cups task shows very similar results, but is notable for having only one return Home (between Movements 2 and 3) and four total movements, defined by velocities of hand (gray) and green and blue cups (green and blue; top plots).

research has shown that participants rarely fixate on objects or areas irrelevant to the goal of a task (Land & Hayhoe, 2001; Hayhoe et al., 2003; Land, 2009; Tatler et al., 2011), we selected specific regions during each phase of movement for analysis. The areas of interest (AOIs) within each phase were defined as the current location being acted on by the hand (*Current*), the

future location that the hand will act upon when it has completed its current action (*Future*), and the hand itself or an object being moved by the hand when no other AOI is being fixated (*Hand in Flight*). Given that each task had several discrete phases for each object movement, Current and Future AOIs were not static but were specifically assigned to each phase, as outlined

AOI	Distance tolerance (cm)
Pasta box transfer task fixation tolerances	
Hand	0.00
Neutral	2.50
Home	5.00
Mid Shelf Target	10.00
High Shelf Target	10.00
Start/End Target	22.50
Pasta box	5.00
Cup transfer task fixation tolerances	
Hand	0.00
Neutral	2.50
Home	1.00
Near Target 1	5.00
Near Target 2	5.00
Far Target 1	7.50
Far Target 2	7.50
Green cup	3.00
Blue cup	3.00

Table 1. The minimum distance between the gaze vector and each area of interest (AOI) required to constitute a fixation.

in the Supplementary Material S1. Additionally, because the participant's hand is indistinguishable from the Current AOI during the Grasp and Release phases, it is included in the Current AOI during these phases. Since we had markers on the moving objects, the hand, and the task apparatus (e.g., the table), we were able to represent each AOI as a physical object in the virtual space created in the motion-tracking coordinate frame. A fixation to an AOI was said to occur when the distance between gaze vector and AOI was sufficiently small (see Table 1) and the velocity from gaze vector to AOI was also sufficiently low (0.5 m/s). To account for blinks, any brief periods (<100 ms) of missing data in each AOI fixation were filled in. Then, to avoid erroneous fixation detection (e.g., fly-throughs), any brief fixations (<100 ms) were removed. A full list of the minimum distances (from intersection or 0 cm to 22.5 cm) between the gaze vector and the bounding box of each AOI that constituted a fixation is shown in Table 1.

Dependent measures

Given the specific objective to derive general principles governing eye-movement behavior during object-related actions, we measured the duration of each phase (Reach, Grasp, Transport, Release), the number of fixations to the Current and Future AOIs in each phase and to the Hand in Flight AOI during the Reach and Transport phases, and the percent fixation time to the Current and Future AOIs in each phase and

to the Hand in Flight AOI during the Reach and Transport phases. In addition to these three phase-specific dependent measures, we calculated a series of measures quantifying the latency of the eye arriving and leaving the site of an object Pick-up (at Transport start) and Drop-off (at Transport end). These four latency measures were calculated for each object movement in both tasks (three movements for the Pasta task, four for the Cups task). The definitions for each measure are given in the following subsections.

Duration

The time in seconds spent in each phase as determined by our segmentation.

Number of fixations

The number of distinct continuous (>100 ms) fixations to an AOI in a given phase.

Percent fixation time

The amount of time fixated on an AOI in a phase divided by the total duration of that phase, multiplied by 100. Note that the results presented here are averages of the trials, for each participant, where a fixation occurred. Trials without fixations were not included in this average.

Eye-arrival latency at Pick-up and Drop-off

Eye-arrival latency (EAL) at Pick-up was defined as Transport start time minus the time of eye arrival at the Pick-up location. EAL values at Pick-up were positive if the eyes began fixating on the object before Transport began, and negative if the eyes began fixating on the object after Transport began. EAL at Drop-off was defined as Transport end time minus the time of the eye arriving at the Drop-off location. EAL values at Drop-off were positive if the eyes began fixating on the target before Transport ended, and negative if the eyes began fixating on the target after Transport ended. See Figure 3 for a visual description.

Eye-leaving latency at Pick-up and Drop-off

Eye-leaving latency (ELL) at Pick-up was defined as Transport start time minus the time of the eye leaving the Pick-up location or object. ELL values at Pick-up were positive if the eyes ended their fixation on the object before Transport began, and negative if the eyes ended their fixation on the object after Transport began. ELL at Drop-off was defined as Transport end time minus the time of the eye leaving the Drop-off location. ELL values at Drop-off were positive if the

eyes ended their fixation on the target before Transport ended, and negative if the eyes ended their fixation on the target after Transport ended. See Figure 3 for a visual description.

Overview of statistical analysis

For each participant, each of the dependent measures was calculated for every trial, then averaged across trials. For duration, number of fixations, and percent fixation time to Current and Future AOIs, each participant had one value for each combination of task (Pasta and Cups), movement (three in Pasta, four in Cups), and phase (Reach, Grasp, Transport, and Release). For number of fixations and percent fixation time to the Hand in Flight AOI, each participant had one value for each combination of task (Pasta and Cups), movement (three in Pasta, four in Cups), and Reach or Transport phase. For the latency measures (EAL and ELL at Pick-up and Drop-off), each participant had four values for each object movement. For the specific repeated-measures analyses of variance (RMANOVAs) described in the following section, significant main effects or interactions were reported if the Greenhouse–Geisser-corrected p value was less than 0.05. Significant interactions were followed up with simple main-effect single-factor ANOVAs. Post hoc tests were run for significant main effects and simple main effects comparing all possible pairwise comparisons of the relevant factor, and were conducted using a Bonferroni correction with a corrected $p < 0.05$ marking a significant effect.

Three analyses were conducted. First, we compared the Pasta and Cups tasks to look for task similarities and differences ($n = 16$ for this comparison, since three Pasta data sets and one Cups data set were removed, as described earlier). In this comparison, because the two tasks had a different number of movements, the measure values were collapsed across movement to yield one value per phase per task, yielding a 2 (Task) \times 4 (Phase) RMANOVA design applied to the duration, Current, and Future measures, and a 2 (Task) \times 2 (Phase) RMANOVA design applied to the Hand in Flight and latency measures. Then, to understand the specific eye-gaze patterns within each task, each of the Pasta ($n = 17$) and Cups ($n = 19$) tasks was analyzed individually. This resulted in: a 3 (Movement) \times 4 (Phase) RMANOVA design for the Pasta task duration, Current, and Future measures; a 3 (Movement) \times 2 (Phase) RMANOVA design for the Pasta task Hand in Flight and latency measures; a 4 (Movement) \times 4 (Phase) RMANOVA design for the Cups task duration, Current, and Future measures; and a 4 (Movement) \times 2 (Phase) RMANOVA design for the Cups task Hand in Flight and latency measures.

Results

General normative eye behavior during sequential object movement

Here, we present the similarities of the eye-behavior measures during sequential object movements between the two tasks. Summarized statistical results can be found in Tables 2–4, containing (respectively) a summary of the comparison between Pasta and Cups tasks, the Pasta task analysis, and the Cups task analysis. Graphical visualization of percent fixation time, number of looks per phase, EAL/ELL measures is shown in Figure 4.

Predictable durations of phases of movement

When we collapsed across movements in order to compare the two tasks, the average duration of each phase was more similar than we expected. Across tasks (Table 2), average Reach-phase durations were not significantly different (Pasta = 0.62 s, Cups = 0.64 s), nor were average Grasp-phase (Pasta = 0.20 s, Cups = 0.19 s) or average Release-phase (Pasta = 0.32 s, Cups = 0.29 s) durations. Average duration values for Transport phases between the two tasks were significantly different (Pasta = 1.17 s, Cups = 1.10 s), but quite close in magnitude. While we take due caution in interpreting a main effect in light of an interaction (see Table 2), the main effect of phase ($F = 1,230.56$, $p = 2.48 \times 10^{-33}$) was interesting in that the Transport phases in both tasks were significantly longer than all other phases ($p < 2.00 \times 10^{-16}$). This prolonged movement with the object in hand and during Drop-off was also reflected in changes in eye-movement behavior that are discussed later.

Notably, the duration of the first Reach (i.e., Movement 1) in the Pasta task was disproportionately longer than subsequent Reaches, even though the hand traveled a relatively short distance. In the Pasta task, this might have been due to a need to turn the body, but this held for the Cups task as well, which did not require a body turn (Table 4). In the Cups task, other than the third Reach (which traverses a larger distance), the first movement again stood out as longest. We attributed these high duration values for Reach phases at the beginning of both tasks to slower movement times resulting from inertia, although they could also possibly have been due to the complexity of movement planning. That is, at the beginning of both tasks, participants began from a stationary position and had to compute movement information about the entire movement sequence. Thus, as recent work is now suggesting, some of this computation, or cognition, could have leaked into the first movement, resulting in it taking longer (Song & Nakayama, 2009; Gallivan &

	Duration (s)		Interaction** task × phase	
	Pasta	Cups	F	Pairwise
Reach	0.62	0.64	Ns	
Grasp	0.20	0.19	Ns	
Transport	1.17	1.10	**	Cups ≪ Pasta
Release	0.32	0.29	Ns	
	Fixation number to Current (#)		Phase effect*	
	Pasta	Cups		
Reach	1.01	1.02		
Grasp	0.95	0.94		Phase pairwise: Grasp < Reach
Transport	1.02	1.02		
Release	0.99	0.98		
	Fixation time to Current (%)		Interaction** task × phase	
	Pasta	Cups	F	Pairwise
Reach	60.75	78.22	**	Pasta ≪ Cups
Grasp	84.80	78.77	ns	
Transport	66.25	75.18	**	Pasta ≪ Cups
Release	70.19	68.75	ns	
	Fixation number to Future (#)		Interaction** task × phase	
	Pasta	Cups	F	Pairwise
Reach	0.02	0.01	ns	
Grasp	0.04	0.05	ns	
Transport	0.04	0.03	ns	
Release	0.10	0.27	**	Pasta ≪ Cups
	Fixation time to Future (%)		Interaction** task × phase	
	Pasta	Cups	F	Pairwise
Reach	0.20	0.12	ns	
Grasp	1.39	1.74	ns	
Transport	0.35	0.29	ns	
Release	1.79	8.15	**	Pasta ≪ Cups
	Fixation number to Hand in Flight (#)		Task effect*, phase effect**	
	Pasta	Cups		
Reach	0.02	0.08		Task pairwise: Pasta ≪ Cups
Transport	0.63	0.70		Phase pairwise: Reach ≪ Transport

	Fixation time to Hand in Flight (%)		Phase effect**	
	Pasta	Cups		
Reach	0.14	1.25		Phase pairwise: Reach ≪ Transport
Transport	7.04	7.85		
	Eye-arrival latency (s)		Interaction* task × phase	
	Pasta	Cups	F	Pairwise
Pick-up	0.57	0.70	**	Pasta ≪ Cups
Drop-off	0.77	0.83	**	Pasta ≪ Cups
	Eye-leaving latency (s)		Phase effect**	
	Pasta	Cups		
Pick-up	−0.06	−0.06		
Drop-off	−0.22	−0.19		Phase pairwise: Drop-off ≪ Pick-up

Table 2. Pasta box (Pasta) and Cup (Cups) transfer task mean dependent measures with one- and two-way significant effects. Notes: ns = no significant main or interaction effect; * and ** appearing in table headings (e.g., Interaction **) indicate the significance of the interaction or main effect listed; * and ** appearing in the F column indicate the significance of simple main effects; **p* < 0.05; ***p* < 0.005. For significant pairwise contrasts and directions, < indicates *p* < 0.05; ≪ indicates *p* < 0.005.

Chapman, 2014). By comparison, subsequent movements started from a hand that was already moving and could have their planning completed during the first (or earlier) movements, and thus did not have any resulting planning or momentum delays.

Participants fixate temporally relevant objects and areas

Participants spent the majority of time fixating on objects and areas that were currently being acted upon by the hand or would be acted upon by the hand in the immediate future. When the percent fixation times for each of the Current, Future, and Hand in Flight AOIs (Tables 3 and 4, Figure 4) were added together and then averaged for each task, participants spent 73.2% of the Pasta task fixating on these three areas combined and 80.1% of the Cups task. Of the remaining task time, 8.6% of the Pasta task and 6.9% of the Cups task had no eye data, likely due to blinks. Much of the remaining 15%–20% of time spent not fixating on the Current, Future, or Hand in Flight AOIs was likely taken up by saccades and head turns where the eye had no discernible fixation. In line with previous literature,

	Duration (s)			Interaction** movement × phase	
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	0.69	0.53	0.66	**	2 << 1,3
Grasp	0.28	0.15	0.18	**	2 << 1,3; 3 << 1
Transport	1.07	1.16	1.27	**	1 << 2,3; 2 << 3
Release	0.29	0.28	0.39	**	1 << 3; 2 << 3
Fixation number to Current (#)			Movement effect*, Phase effect**		
	Movement 1	Movement 2	Movement 3		
Reach	1.00	1.00	1.02	Movement pairwise: 1 < 3	
Grasp	0.97	0.91	0.97	Phase pairwise: Grasp < Reach, Grasp < Transport	
Transport	1.01	1.01	1.04	Release < Transport	
Release	0.98	0.98	1.01		
Fixation time to Current (%)			Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	37.30	75.58	66.39	**	1 << 2,3; 3 << 2
Grasp	80.25	85.72	90.89	ns	
Transport	73.09	75.57	48.34	**	3 << 1,2
Release	71.15	79.65	62.78	**	1 < 2; 3 << 2
Fixation number to Future (#)			Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	0.04	0.01	0.00	ns	
Grasp	0.03	0.07	0.00	ns	
Transport	0.00	0.00	0.12	**	1 < 3; 2 < 3
Release	0.08	0.00	0.20	**	2 < 3
Fixation time to Future (%)			Interaction* movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	0.54	0.11	0.00	ns	
Grasp	0.33	3.59	0.00	ns	
Transport	0.00	0.00	1.08	**	1 < 3; 2 < 3
Release	1.36	0.00	3.76	*	ns, trend 2 < 3
Fixation number to Hand in Flight (#)			Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	0.00	0.06	0.08	ns	
Transport	0.34	0.77	0.83	**	1 << 2,3
Fixation time to Hand in Flight (%)			Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Reach	0.03	0.61	1.15	ns	
Transport	2.14	10.75	9.71	**	1 << 2,3
Eye-arrival latency (s)			Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Pick-up	0.53	0.55	0.62	*	1 < 3; 2 << 3
Drop-off	0.78	0.87	0.61	**	3 << 1,2
Eye-leaving latency (s)			Interaction* movement × phase		
	Movement 1	Movement 2	Movement 3	F	Pairwise
Pick-up	0.03	-0.11	-0.11	**	2 << 1; 3 << 1
Drop-off	-0.20	-0.22	-0.25	*	3 < 1

Table 3. The means of the dependent measures calculated for the pasta-box transfer task, with one- and two-way effects. Notes: ns = not significant; **p* < 0.05; ***p* < 0.005. For significant pairwise contrasts and ** directions, < indicates *p* < 0.05; << indicates *p* < 0.005.

	Duration (s)				Interaction** movement × phase	
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Reach	0.68	0.52	0.88	0.48	**	4 << 1,3; 4 < 2; 2 << 1,3; 1 << 3
Grasp	0.17	0.22	0.22	0.14	**	4 << 1,2,3; 1 << 3; 1 < 2
Transport	1.02	1.16	1.16	1.04	**	1 << 2,3; 4 << 2,3
Release	0.30	0.31	0.30	0.32	ns	
Fixation number to Current (#)				Phase effect*		
	Movement 1	Movement 2	Movement 3	Movement 4		
Reach	1.07	1.01	1.03	0.98	Phase pairwise: Grasp < Reach Grasp < Transport	
Grasp	0.93	0.95	0.94	0.89		
Transport	1.02	1.01	1.03	1.03		
Release	0.96	0.99	0.97	1.00		
Fixation time to Current (%)				Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Reach	67.11	92.25	76.82	79.37	**	1 << 2; 1 < 3; 3 < 2
Grasp	76.13	83.11	86.02	69.99	ns	
Transport	78.66	80.09	73.94	65.25	**	4 << 2; 4 < 1; 3 < 2
Release	48.88	82.33	64.37	79.79	**	1 << 2,3,4; 3 << 2; 3 < 4
Fixation number to Future (#)				Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Reach	0.00	0.01	0.03	0.00	*	ns, trend 3 larger
Grasp	0.07	0.04	0.05	0.03	ns	
Transport	0.08	0.00	0.01	0.01	*	ns, trend 1 larger
Release	0.58	0.00	0.43	0.18	**	2 << 1,3; 3 << 1; 4 << 1; 4 < 3
Fixation time to Future (%)				Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Reach	0.00	0.10	0.26	0.03	ns	
Grasp	1.94	1.27	1.69	1.30	ns	
Transport	0.72	0.00	0.06	0.20	ns	
Release	17.66	0.00	10.87	4.29	**	2 << 1,3; 2 < 4; 4 << 1; 4 < 3
Fixation number to Hand in Flight (#)				Movement effect**, phase effect**		
	Movement 1	Movement 2	Movement 3	Movement 4		
Reach	0.05	0.05	0.10	0.14	Movement pairwise: 1 << 3; 1 < 4; 2 < 3 Phase pairwise: Reach << Transport	
Transport	0.55	0.63	0.85	0.78		
Fixation time to Hand in Flight (%)				Movement effect**, phase effect*		
	Movement 1	Movement 2	Movement 3	Movement 4		
Reach	0.56	0.35	0.92	2.95	Movement pairwise: 1 << 3 Phase pairwise: Reach << Transport	
Transport	5.16	7.12	10.24	10.07		
Eye-arrival latency (s)				Interaction** movement × phase		
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Pick-up	0.66	0.71	0.90	0.55	**	4 << 2,3; 4 < 1; 1 << 3; 2 << 3
Drop-off	0.80	0.92	0.86	0.69	**	4 << 2,3; 4 < 1; 1 << 2
Eye-leaving latency (s)				Interaction* movement × phase		
	Movement 1	Movement 2	Movement 3	Movement 4	F	Pairwise
Pick-up	-0.01	-0.05	-0.11	-0.11	*	4 < 1; 4 < 3
Drop-off	-0.14	-0.25	-0.18	-0.25	**	2 << 1,3; 4 << 1, 4 < 3; 3 < 1

Table 4. The means of the dependent measures calculated for the cup transfer task, with one- and two-way effects. Notes: ns = not significant; * $p < 0.05$; ** $p < 0.005$. * & ** appearing in table headings (e.g., Interaction **) indicate the significance of the interaction or main effect listed. * & ** appearing in the F column indicate the significance of simple main effects. For significant pairwise contrasts and directions, < indicates $p < 0.05$; << indicates $p < 0.005$.

this is further evidence that during goal-directed tasks, visual attention is dedicated mostly not just to relevant areas and objects but to those objects or areas that are most temporally relevant to the specific subset of the task being conducted (Land & Hayhoe, 2001; Hayhoe et al., 2003; Land, 2009; Tatler et al., 2011).

Temporal patterns of visual fixations when interacting with objects

When interacting with objects, participants exhibited consistent temporal eye-movement patterns. The number of fixations to the Current AOI in all phases of both tasks was, on average, 1 (Tables 2–4, Figure 4). This means that participants chose to hold their fixation at one location before moving to the next area, instead of shifting fixation back and forth between areas. As well, participants' eyes arrived early at the object to be picked up during Reach, as shown by the EAL values during Pick-up for both tasks (Tables 3 and 4, Figure 4) exceeding 0.5 s. Participants ended this fixation at approximately the same time that the object began moving, as shown by the ELL values during Pick-up for both tasks (Tables 3 and 4, Figure 4) clustering near 0 s. However, these values were almost all slightly negative, indicating that the object began moving before the eyes ended their fixation, suggesting that some degree of confidence in grip was required before the eyes left an object. This is consistent with previous literature stating that vision is released when another sensory modality (touch) takes over (Land & Hayhoe, 2001; Hayhoe et al., 2003; Land, 2009; Tatler et al., 2011). After leaving the Pick-up location, the eyes moved to the Drop-off location (EAL Drop-off), again leading the hand by an average of more than 0.6 s. Finally, there was consistency across tasks in the relative eye-fixation behavior during Pick-up and Drop-off events. Specifically, the ELL values in both the Pasta and Cups tasks (Tables 3 and 4, respectively) at Pick-up were, on average, more than 100 ms shorter than those at Drop-off, suggesting that the longer Release durations already reported were also accompanied by an eye that lingered on just-released objects longer than just-grasped ones.

Eyes rarely fixate on the hand, but fixate on the object in hand during early Transport

The number of fixations and percent fixation time to a participants' Hand in Flight during Reach were extremely low (Tables 2–4, Figure 4). On average, a fixation to the Hand in Flight during Reach occurred less than once across all 20 trials, accounting for approximately 1% of Reach duration (Table 2). Importantly, during Transport we saw an increase in the number of fixations and percent fixation time to the

Hand in Flight, with more than one fixation every two trials, accounting for more than 7% of Transport duration (Table 2). These magnitudes were still comparatively low, but they suggest that participants took extra time fixating on the object as it was being lifted at the beginning of the Transport phase. This was further supported by the EAL values during Pick-up (e.g., small negative values) for both tasks already discussed.

Biomechanics of movement affect eye-movement behavior

Variations in biomechanics caused differences in eye-movement behavior, which was most evident in the results from our analysis of the Pasta task (Table 3, Figure 4a). Postural constraints have previously been found to affect eye behavior; specifically, Di Cesare et al. (2013) have found that eye movements were delayed when whole-body rotations were used to fixate a target. As well, Freedman (2008) has shown that saccades increased in velocity as head movements were restricted. When participants were interacting with the High Shelf Target of the Pasta task, which required their arm to cross their body's midline at or above shoulder height, they were restricted from turning their head or neck back toward the Future target location (either the Home position or the Start/End Target). This resulted in a percent fixation time to Current during the Release of Movement 2 that was larger than those of the other two movements. Also, the EAL during the Drop-off of Movement 3 was much shorter than those of the other two movements, because the head could not turn toward the Start/End Target since the arm was grasping high and across the body. By comparison, the ELL during the Pick-up of Movement 1 was shorter than those of the other two Movements, likely due to the ease with which participants could turn their head away from the Start/End Target while still interacting with the Pasta box.

Objects and areas outside the field of view were fixated less

The location of the Start/End Target of the Pasta task elicited intriguing behaviors. The requirement to turn and fixate an AOI outside the initial field of view resulted in lower percent fixation time to Current (Table 3, Figure 4a), as seen in the Reach of Movement 1 and the Transport of Movement 3. This was also reflected in a complete absence of fixations to Future during the Reach and Grasp of Movement 3—it was simply too far away and out of sight. The EAL of the Drop-off of Movement 3 was significantly shorter than those of the other two movements, meaning that participants' eyes were not fixating on this location as

early in the movement, due to this area being outside the field of view. By comparison, in the Cups task all objects and areas were quite easily fixated by participants, requiring little turning of the head and no turning of the body. This difference in task requirements accounted for the largest differences in the comparison of Pasta versus Cups tasks (Table 2). The aforementioned fixation reductions to the Start/End Target in the Pasta task resulted in significant task differences, with the Pasta task showing an overall reduction in percent fixation time to Current during Reach and Grasp as well as shorter EAL for both Pick-up and Drop-off.

The goal of the next movement dictates visual behavior of the current Release

In our study, there were several distinct hand actions: reaching toward objects, grasping objects, transporting objects, releasing objects, and touching the Home position. We found that participants required more visual attention in advance if the hand was grasping an object than if it was merely touching an area, which is in accordance with previous literature (Hayhoe et al., 2003). To highlight this, we looked at the differences in eye-movement behavior when the hand was releasing an object where the next action was to touch the Home position as compared to where the next action was to reach and grasp another object. The Release straight to Reach/Grasp occurred only in the Cups task (Table 4, Figure 4b)—specifically, the Release of Movement 1 led directly to the Reach of a cup for Movement 2, whereas the Release of Movement 2 was followed immediately by a return of the hand to touch the Home position. A similar difference occurred between the Releases of Movements 3 (to cup) and 4 (to Home). This comparison revealed a greater number of fixations to Future and higher percent fixation time to Future for the Release phase of Movements 1 (fixations = 0.58; percentage = 17.66) and 3 (fixations = 0.43; percentage = 10.87) than for Movements 2 (fixations = 0.00; percentage = 0.00) and 4 (fixations = 0.18; percentage = 4.29). Correspondingly, the percent fixation time to Current dropped during the Release of Movements 1 (48.88%) and 3 (64.37%) compared to Movements 2 (82.33%) and 4 (79.79%). These results, visualized in Figure 4b, show that participants stopped fixating the Drop-off location earlier in the Release phases when they were reaching for the next cup as compared to when they were releasing to reach toward the Home position. This finding was also corroborated in the latency measures, where the ELL during Drop-off in Movements 1 (−0.14 s) and 3 (−0.18 s) was about 100 ms shorter in Movements 2 (−0.25 s) and 4 (−0.25 s). A possible explanation is that when the next relevant movement was an object interaction, there was greater

urgency for the eyes to leave the Current location, sacrificing a small amount of accuracy in the Release to fixate on the next object to be interacted with for a more reliable next Reach and Grasp.

Discussion

We examined the eye-movement behavior of participants as they moved everyday objects in simulated real-world environments, with specific required movement sequences. By integrating eye and motion tracking, this study provides temporal accuracy to quantifying eye-movement behavior during real object interactions beyond the work of previous studies (Land et al., 1999; Land & Hayhoe, 2001; Hayhoe et al., 2003). Across tasks and across movements, phases (Reach, Grasp, Transport, and Release) were relatively consistent in duration. Participants spent the majority of task time fixating on areas relevant to the task, and visual fixation tended to precede object interaction by more than half a second. Little fixation was dedicated to a participant's own hand except in the early moments of the Transport phase. The physical setup of the task and resulting body-movement requirements affected visual fixation patterns such that objects outside the field of view were fixated less, specific postures (e.g., arm across the body) reduced eye lead times, and objects to be grasped required earlier visual feedback compared to areas not requiring a grasp.

Due to the 60-Hz frame rate of the eye tracker used in this study, we were not able to accurately capture short saccades or saccade dynamics, and any measures calculated from the fixation data could have been off by as much as 0.017 s. However, this eye tracker provided a data set that was robust enough for our study, as we were seeking to describe general visual-behavior patterns rather than finer-detail patterns like microsaccades. Furthermore, the precision afforded by eye and motion tracking was augmented by our experimental design, which allowed analysis of 40 trials (20 of each task) of short-duration (~10 s) tasks from 20 different participants; this is in contrast to previous work, where often only a single trial across a handful of participants has been analyzed (Land et al., 1999; Land & Hayhoe, 2001; Hayhoe et al., 2003), or even only a subset of trials (Parr, Vine, Harrison, & Wood, 2017). Although it has been shown that visual behavior changes as individuals carry out repeated trials of visuomotor tasks (Sailer, Flanagan, & Johansson, 2005; Foerster, Carbone, Koesling, & Schneider, 2011), all participants in our study carried out a similar number of trials and were given an opportunity to practice the task before recording trials began. That is, while there may have been practice effects, that was not

our primary factor of interest, and experience was held constant across participants. We hypothesize that the same general pattern of fixations would hold for all trials in the data set, and that these patterns would hold for people carrying out well-practiced everyday tasks like these in their own homes. In other words, we hypothesize that any practice effects, while likely to be present, would cause minimal changes to the general fixation patterns we report. What enabled us to analyze this comparatively large data set and thereby increase our statistical power was the development of a custom data-analysis tool that provides a visualization of synchronized eye and motion data, separates eye-movement data into phases based on hand and object velocity and grip aperture, and automates the detection of fixations on AOIs and the calculation of eye-movement measures.

Our results strengthen many of the findings reported previously regarding eye-movement behavior during object-interaction tasks in the real world (Land & Hayhoe, 2001; Hayhoe et al., 2003; Land, 2009; Tatler et al., 2011) and fit well with previous literature stating that task and context affect visual gaze behavior (Rothkopf, Ballard, & Hayhoe, 2007). Land and Hayhoe (2001) found that a participant's own hands are very rarely fixated. Our data set and analysis techniques allow us to refine this point: Participants rarely fixate their own hand when reaching toward an object but tend to maintain a brief fixation on the object in the hand when beginning to pick up and transport the object. Similarly, we were able to precisely quantify the pattern and timing of the eye leading the hand. During a tea-making task, it has been reported that the eyes led the hand by an average of 0.56 s (Land et al., 1999; Land & Hayhoe, 2001). In our study, this average value (for arrival time before Pick-up) ranged from 0.53 to 0.90 s depending on the specifics (e.g., height, distance) of the movement. The relatively large range arrived at in our more detailed analysis supports a *minimum* fixation time before hand-object interaction to be about half a second, while allowing for an advance fixation to be much longer if the demand for visual information from another location is not sufficient to drive the eyes away. When the movement allows, this time can be increased, and in cases where this minimum is not attainable (e.g., fast movements or reaches toward objects outside of the current field of view), we predict consequences to the accuracy of the subsequent interaction.

In addition to providing precise descriptive information about eye-movement behavior during object interactions, the current study fits within current theoretical frameworks. Specifically, we propose a connection to Baldauf and Deubel's (2010) Attentional Landscapes Theory of Visual Attention. That theory proposes that, during movement planning, covert

attention is automatically distributed to upcoming action locations that are visible in the current field of view, forming a landscape of attention with preferential processing (e.g., hills) at intended action sites and less or suppressed attention (e.g., valleys) at nonaction locations. The peak of the landscape represents the most behaviorally relevant location, and eye movements are driven toward locations of high relevance. This theory is supported by a series of studies where visual targets were more easily discriminated at the upcoming target locations of multistep eye and reach movements (for a summary, see Baldauf & Deubel, 2010). In a 2006 study, Baldauf, Wolf, and Deubel instructed participants to keep visual fixation on a central cross while making reaches to two or three targets in a specific sequence. A secondary letter-discrimination task presented briefly just prior to the first reach was used to assess covert visual attention at several locations, including each of the upcoming reach targets. It was found that participants were far more successful at discriminating the letters flashed at the areas of the upcoming movement goals compared to areas that were irrelevant to the upcoming task. Interestingly, discrimination for letters presented before the first movement started was improved not just at the first movement goal but also at the second and third movement goals (though performance diminished with increased sequential positioning). This is strong evidence that the covert visual-attention system dedicates more resources to locations of future movements. Baldauf and Deubel (2008) strengthened this theory by finding similar results in electrophysiological data. It was shown that the N1 event-related potential found using electroencephalography is enhanced when a dot is flashed at a location of an upcoming movement goal. This shows a neural correlate for the previous behavioral results that covert attention shifts to locations of upcoming goal-directed movements. These previous studies strongly suggest that covert visual attention is allocated to locations of future movements and primes actors to fixate these relevant locations as they carry out the task.

This obligatory allocation of attention to upcoming action sites, which in turn facilitates and drives eye movements to those locations, provides a nice framework to understand the pattern of results observed in the current study. For our tasks, the Attentional Landscapes framework hypothesizes that participants devote the most attention toward the immediate action target (e.g., a cup they are reaching toward) while still dedicating some attention to the next target of action (e.g., where the cup will be put down). As the success of the first movement goal becomes more certain (e.g., the hand-to-object distance is sufficiently small, or proprioceptive feedback is received during a grasp or release), the landscape begins to shift, with the next target of

action receiving an increasing amount of attention. At some point, near the onset of object movement, the scales are tipped and the eyes are driven away from the site of current action toward the site of future action. Specific task demands (object location and size, grasp type, posture, etc.) shape the dynamic landscape. One notable example in our data are hand-to-Home movements: This movement goal has limited precision requirements and thus rarely demands sufficient attention to drive eye movements to its location. It should be noted, however, that by restricting our analysis to eye-movement patterns during natural behavior, we deprived ourselves of secondary measures (such as the discrimination task of Baldauf et al., 2006, or the EEG of Baldauf & Deubel, 2008) to confirm our speculation of parallel processing. In fact, our results show that participants behave in a consistent and predictable serial manner. However, we feel that these results are still best explained by parallel processing. In normal, everyday object interactions, we envision an attentional landscape that rises and falls at different locations in our environment based on the current task demands and information available. For example, the landscape will be high at an object you are reaching toward, driving the eyes to that location. However, as Baldauf and colleagues have shown, as you reach, another hill of activity will be forming at the next location of action, namely, where the object will be placed. As your approaching hand begins receiving haptic information signaling a successful grasp, the landscape peak at the location of current interaction will diminish, while the peak at the drop-off site will grow. When these peaks shift in prominence (e.g., the drop-off peak becomes the biggest), the eyes are now driven away from the site of interaction toward the drop-off location.

While successful at describing the majority of the observed eye movements in our task, the Attentional Landscapes theory has no explicit mechanism for allocating attention to objects that are not currently in the field of view. Since several of the interesting findings reported here are caused by reaches toward the side table location in the Pasta task, which is usually out of view, here we offer an important addition to the Attentional Landscapes framework to account for the observed data. To fill in the missing information from a landscape with no real-time visual information about an out-of-sight movement goal, we appeal to the two-visual-stream hypothesis (for a review, see Goodale, 2011). This theory argues that there are anatomical and functional differences between visual information that flows dorsally from the primary visual cortex (vision for action) and visual information that flows ventrally from the primary visual cortex (vision for perception). Classically, the dorsal vision-for-action stream relies on real-time visual information—and thus is the prime

candidate to sculpt the Attentional Landscape for objects within view. In fact, it has been shown in monkeys that the posterior parietal cortex, a major sensory area in the dorsal stream, encodes for multiple sequential reaching goals before a reach has been initiated (Baldauf, Cui, & Andersen, 2008). However, with objects outside the field of view, we believe participants must rely on the ventral vision-for-perception stream. This shift to more “offline” information comes at a cost, as has been reported previously (Hu, Eagleson, & Goodale, 1999), as we find that participants spend more time grasping when reaching toward an object at this out-of-sight location. Previous work has shown that grip apertures are larger when immediate visual feedback is not available during movement (Berthier, Clifton, Gullapalli, McCall, & Robin, 1996; Hu et al., 1999), and this is consistent with our finding that the distance and time participants spent grasping (e.g., closing finger and thumb) were increased for the side table location in the Pasta task. This theory aligns with previous literature positing that implicit memory representations of the spatial structure of the environment are built for later use during object interactions (Chun & Nakayama, 2000; Hayhoe et al., 2003). It also aligns with more recent work in functional magnetic resonance imaging showing that areas in the ventral stream combine discrete images of a panoramic scene, both in and out of view, to create a common representational space used to help form a continuous visual experience for the individual (Robertson, Hermann, Mynick, Kravitz, & Kanwisher, 2016). As well, previous research with functional magnetic resonance imaging has shown that areas in the ventral stream known for visual object recognition are re-recruited and reactivate V1 before a reach and grasp is initiated after a delay period in darkness (Singhal, Monaco, Kaufman, & Culham, 2013). All this supports the theory that the ventral stream is at least partly responsible for maintaining a representation of objects that are outside the current field of view, thereby allowing for more effective interaction toward objects that are nearby but not currently visible.

It is important to note that our study does not provide any direct insight into the role that peripheral vision plays in visually guided object interactions, as only the central fixation point was recorded. It has been shown not only that different cortical systems are dedicated to reaching toward centrally as opposed to peripherally located targets (Prado et al., 2005) but also that these different cortical systems each supply different, but necessary, characteristics of a successful grasp and transport of an object; peripheral vision supplies a wide field of view, providing environmental information aiding reaching and transporting, while central vision supplies high-resolution information aiding grasping and transporting (Sivak & MacKenzie,

1990). Thus, while peripheral vision was almost certainly involved in the early object targeting involved in our task (e.g., during the Reach phase), our eye-tracking technique does not allow us to quantify this contribution.

The integration of the two-visual-stream hypothesis with the Attentional Landscapes theory explains how the current environment, both within and outside of the field of view, could be represented in the brain, and why the eyes are primarily driven to each subsequent action location (and very little else) in our tasks. One critical question that remains to be answered is what drives the dynamics of the landscape. That is, if we take eye movements as the expression of the most relevant locations, then what drives the shift in relevance? Earlier we mentioned that relevance shifts are likely driven by action confidence, a property that must be estimated by the actor from visual and proprioceptive feedback. This suggests that manipulating the feedback available to actors should fundamentally alter eye gaze. Recent studies with users of prosthetic limbs (Sobuh et al., 2014) and participants controlling artificial hands (Parr et al., 2017) confirm that this is the case—with reduced proprioceptive and tactile feedback, fixations to the hand and object in flight are massively increased. Put into the theoretical framework, these prolonged fixations to the hand and object during interaction arise because without sensory feedback, a lack of confidence in the object manipulation essentially freezes the landscape, leaving the object as the most relevant location and preventing the eyes from moving ahead to the drop-off site. As well, it has recently been shown that individuals experiencing neurodegenerative brain disorders like Alzheimer's disease and Parkinson's disease exhibit characteristic visuomotor profiles that could be used in a clinical setting (de Boer, van der Steen, Mattace-Raso, Boon, & Pel, 2016). An extension of our study to clinical populations like those who use upper-limb prostheses and those with movement disorders such as Alzheimer's disease and Parkinson's disease could provide interesting information and insight into altered visuomotor behavior in impaired states.

Conclusions

By combining eye and motion tracking and using a sequential and repeated task design and a novel analysis tool, we were able to provide greater temporal and spatial accuracy in quantifying eye behavior during object interaction. We confirm and extend many of the seminal findings of Land and Hayhoe (2001) and demonstrate how eye behavior can be explained by appealing to an updated Attentional Landscapes model

(Baldauf & Deubel, 2010) that integrates the two-visual-stream hypothesis (Goodale, 2011) to account for movements toward out-of-view objects. Future work will focus on what drives the landscape to shift, and under what conditions (e.g., reduced proprioceptive and tactile feedback) anomalous eye-movement patterns are observed. Finally, a comprehensive description of object interaction should also include kinematic analysis of the moving body during these tasks. Understanding the nuanced connections between specific body movements and the corresponding eye movements presents an exciting future opportunity.

Keywords: visuomotor control, eye–hand coordination, visual attention, object interaction, eye- and motion-tracking synchronization, sequential object movement

Acknowledgments

We thank Elizabeth Crockett, Brody Kalwajtys, Bret Hoehn, Kory Mathewson, John Luu, and Thomas R. Dawson for their contributions to task design, data collection, and data processing.

This work was sponsored by the Defense Advanced Research Projects Agency (DARPA) BTO under the auspices of Dr. Doug Weber and Dr. Al Emondi through the DARPA Contracts Management Office Grant/Contract No. N66001-15-C-4015.

Commercial relationships: none.

Corresponding author: Craig S. Chapman.

Email: c.s.chapman@ualberta.ca.

Address: Faculty of Kinesiology, Sport, and Recreation and Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Alberta, Canada.

References

- Baldauf, D., Cui, H., & Andersen, R. A. (2008). The posterior parietal cortex encodes in parallel both goals for double-reach sequences. *The Journal of Neuroscience*, *28*(40), 10081–10089.
- Baldauf, D., & Deubel, H. (2008). Attentional selection of multiple goal positions before rapid hand movement sequences: An event-related potential study. *Journal of Cognitive Neuroscience*, *21*(1), 18–29, <https://doi.org/10.1162/jocn.2008.21021>.
- Baldauf, D., & Deubel, H. (2010). Attentional land-

- scapes in reaching and grasping. *Vision Research*, 50(11), 999–1013.
- Baldauf, D., Wolf, M., & Deubel, H. (2006). Deployment of visual attention before sequences of goal-directed hand movements. *Vision Research*, 46(26), 4355–4374, <https://doi.org/10.1016/j.visres.2006.08.021>.
- Berthier, N. E., Clifton, R. K., Gullapalli, V., McCall, D. D., & Robin, D. J. (1996). Visual information and object size in the control of reaching. *Journal of Motor Behavior*, 28(3), 187–197.
- Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: Chicago University Press.
- Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, 7(1–3), 65–81.
- de Boer, C., van der Steen, J., Mattace-Raso, F., Boon, A. J., & Pel, J. J. (2016). The effect of neurodegeneration on visuomotor behavior in Alzheimer's disease and Parkinson's disease. *Motor Control*, 20(1), 1–20.
- Desmurget, M., Pélisson, D., Rossetti, Y., & Prablanc, C. (1998). From eye to hand: Planning goal-directed movements. *Neuroscience and Biobehavioral Reviews*, 6(22), 761–788.
- Di Cesare, C. S., Anastasopoulos, D., Bringoux, L., Lee, P. Y., Naushahi, M. J., & Bronstein, A. M. (2013). Influence of postural constraints on eye and head latency during voluntary rotations. *Vision Research*, 78, 1–5.
- Epelboim, J., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, 37(18), 2597–2607.
- Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, 11(7):9, 1–16, <https://doi.org/10.1167/11.7.9>. [PubMed] [Article]
- Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, 190(4), 369–387.
- Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in Neuroscience*, 8(215), 1–19.
- Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, 51(13), 1567–1587.
- Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, 7(1–3), 43–64.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194.
- 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, 49–63, <https://doi.org/10.1167/3.1.6>. [PubMed] [Article]
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Hu, Y., Eagleson, R., & Goodale, M. A. (1999). The effects of delay on the kinematics of grasping. *Experimental Brain Research*, 126(1), 109–116.
- Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, 21(17), 6917–6932.
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3), 317–340.
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13), 1457–1483.
- Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, 26(1), 51–62.
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25), 3559–3565.
- Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, 369(6483), 742–744.
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328.
- Milner, D., & Goodale, M. (2006). *The visual brain in action*. Oxford, UK: Oxford University Press.
- Neggers, S. F., & Bekkering, H. (2000). Ocular gaze is anchored to the target of an ongoing pointing movement. *Journal of Neurophysiology*, 83(2), 639–651.
- Parr, J. V. V., Vine, S. J., Harrison, N. R., & Wood, G. (2017). Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand. *Journal of Motor Behavior*. Advance online publication, <https://doi.org/10.1080/00222895.2017.1363703>.
- Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41(25), 3587–3596.

- Prado, J., Clavagnier, S., Otzenberger, H., Scheiber, C., Kennedy, H., & Perenin, M. T. (2005). Two cortical systems for reaching in central and peripheral vision. *Neuron*, *48*(5), 849–858.
- Richardson, D. C., & Spivey, M. J. (2008). Eye tracking: Research areas and applications. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering*, 2nd ed. (pp. 1033–1042). New York, NY: Informa Healthcare.
- Robertson, C. E., Hermann, K. L., Mynick, A., Kravitz, D. J., & Kanwisher, N. (2016). Neural representations integrate the current field of view with the remembered 360 panorama in scene-selective cortex. *Current Biology*, *26*(18), 2463–2468.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14):16, 1–20, <https://doi.org/10.1167/7.14.16>. [PubMed] [Article]
- Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience*, *25*(39), 8833–8842.
- Singhal, A., Monaco, S., Kaufman, L. D., & Culham, J. C. (2013). Human fMRI reveals that delayed action re-recruits visual perception. *PLoS One*, *8*(9), e73629.
- Sivak, B., & MacKenzie, C. L. (1990). Integration of visual information and motor output in reaching and grasping: The contributions of peripheral and central vision. *Neuropsychologia*, *28*(10), 1095–1116.
- Sobuh, M. M., Kenney, L. P., Galpin, A. J., Thies, S. B., McLaughlin, J., Kulkarni, J., & Kyberd, P. (2014). Visuomotor behaviors when using a myoelectric prosthesis. *Journal of Neuroengineering and Rehabilitation*, *11*(1): 72.
- Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366.
- 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, 1–23, <https://doi.org/10.1167/11.5.5>. [PubMed] [Article]
- Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.