Adaptive Gaze and Hand Coordination while Manipulating and Monitoring the Environment in Parallel

Jolande Fooken[1,2*], Roland S. Johansson[3†], & J. Randall Flanagan[2†]

[1]Department of Psychology and Centre for Cognitive Science, Technical University of Darmstadt, Darmstadt, Germany

[2]Department of Psychology and Centre for Neuroscience Studies, Queen’s University, Kingston, Ontario, Canada

[3]Department of Integrative Medical Biology, Umeå University, Umeå, Sweden

[*] Corresponding author: jolande.fooken@queensu.ca

[†] Equal contribution

Competing Interests

The authors have no competing interests to disclose.

Acknowledgments

This work was supported by a German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) Research Fellowship to JF, the Swedish Research Council Project (grant 22209 to RSJ), the Canadian Institutes for Health Research (grant RGPIN/05944-2019 to JRF), and the Natural Science and Engineering Research Council of Canada (grant 156173 to JRF).

Abstract

Research on eye-hand coordination has focused on action tasks performed in isolation. However, real world action tasks are often performed concurrently with perception tasks that compete for gaze. Here we examine how participants adapt their eye and hand movements when performing an object manipulation task—in which they repeatedly grasped a ball and inserted it into a slot—while simultaneously monitoring a text display to detect probabilistically occurring letter changes. We varied the visuomotor demands of the action task by having participants use either their fingertips or tweezers. We found that fixations allocated to the action task were exclusively directed to the ball and slot, and were more prevalent when using tweezers. The timing of ball and slot fixations were coupled in time with ball grasp and slot entry. On average, gaze shifted away from the landmarks ~400 ms before contact when using fingertips—allowing the use of peripheral vision to direct the hand—and around the contact time when using tweezers—further allowing central vision to guide the hand as it approached the ball or slot. We found that participants controlled the timing of their hand movements, as well as the timing and patterns (sequence of fixations) of their eye movements, to exploit the temporal regularities of the perception task, thereby lowering the probability that a letter change would occur during action task fixations. Our results illustrate that eye-hand coordination can be flexibly and intelligently adapted when simultaneously acting on and perceiving the environment.

Introduction

Gaze fixations, occurring between eye movements, play a vital role in both perceiving the world and in planning and controlling actions (Yarbus 1967; Land 2006; Hayhoe 2017; Kowler 2011). Although research has extensively examined gaze control in action and perception tasks independently, real-life scenarios often demand concurrent performance of visually guided actions and visual perception tasks (Fooken et al., 2023; Land and Furneaux 1997), leading to a competition for gaze resources. For instance, at a dinner party, diners use their gaze to control their manual actions, such as handling objects, while also using their gaze to survey their surroundings and engage in conversations. Under such circumstances, one would expect that gaze would only be directed to the action task when the support of gaze is most important.

To our knowledge, no prior research has delved into the control policies and strategies governing gaze allocation when there are competing demands for gaze from manual actions and environmental monitoring. To address this gap, we designed an experiment where participants simultaneously performed an object manipulation task and a visual monitoring task that required central vision. The manipulation task involved repeatedly grasping a small ball situated on a platform and inserting it into a slot in a vertical tube, from which the ball returned to the platform (Fig. 1A). Participants completed the task using either their fingertips or tweezers, enabling us to manipulate the visuomotor demands. Concurrently, participants were tasked with monitoring a text display for letter changes, which occurred randomly. Successful ball drops were rewarded, while failures to detect letter changes incurred penalties.

The aim of this paper was to test three novel hypotheses concerning how gaze allocation might be optimized during the concurrent execution of an action task and a perception task that compete for gaze. First, we hypothesized that the prevalence and function of gaze fixations directed to the action task would depend on the visuomotor demands, shaped by the end-effector employed and the phase of the task. Second, we hypothesized that, by observing the temporal statistics of relevant visual events in the environment, individuals would predict when events requiring central vision are more or less likely to occur, and use this information to allocate gaze more efficiently between tasks. Third, we hypothesized that individuals would reduce competition between tasks for gaze resources by adjusting the timing of the action task and, consequently, the timing of required action task fixations. The rationale and motivation for these hypotheses is developed below.

Object manipulation tasks involve a sequence of action phases delineated by kinematic or mechanical events, as illustrated in Figure 1 (Johansson and Flanagan 2009; Johansson et al. 2001). When such tasks are performed in isolation, gaze is almost exclusively directed to targets of action—the ball and slot in our ball-drop task—with the function of gaze changing across action phases (Illamperuma and Fooken, 2024). When moving the hand, or object in hand, towards the vicinity of a target—as in the reach and transport phases of our task—individuals typically fixate the target. This fixation enables fast, automatic feedback control mechanisms that use peripheral vision and gaze-related signals to *direct* the hand toward the target (Saunders and Knill 2003, 2004; Goodale et al., 1986). Once the hand gets close to the object, and more slowly approaches it—as in the ball and slot approach phases in our task—gaze may remain on the target, in which case central vision can be used to *guide* the hand, or grasped object, to the target through more deliberate closed-loop feedback control (Johansson et al. 2001; Ballard et al. 1992; Land 2006). Once the hand or object in hand comes into contact with the target—as in the ball and slot contact phases in our task—tactile feedback becomes available. If gaze still remains on the target, central vision can be used to visually check successful contact. A key question addressed in the current study is which of these functions of gaze gets prioritized when there is competition for gaze.

We expected that when the ball-drop is performed concurrently with the letter change monitoring task, action task fixations would still be directed to the ball and slot. With respect to our first hypothesis, we predicted that when using the fingertips, action task fixations, if observed, would primarily serve the purpose of *directing* the hand via peripheral vision. We predicted that fixations involved in *guiding* the hand via central vision would not be required when using fingertips because, once the hand (or ball in hand) is in proximity to the target, haptic feedback will be used to correct for positioning errors. In contrast, we predicted that when using tweezers, gaze fixations, in addition to being involved in directing the hand, would also be involved in guiding the hand, because tweezers require greater spatial precision, particularly in grasping the ball, and offer limited tactile feedback about the contact state. Note that impaired tactile sensibility is known to increase reliance on visual feedback for object manipulation (Brink and Mackel, 1987; Chemnitz et al., 2013; Jenmalm and Johansson, 1997; Jerosch-Herold, 1993).

Previous work has shown that when concurrently monitoring two locations to detect probabilistic events, individuals can optimize their gaze allocation by learning the temporal regularities of the events at each location and adjusting their gaze accordingly (Hoppe and Rothkopf 2016). This raises the question of whether people can similarly learn and exploit temporal regularities of events when concurrently engaged in a visual monitoring task and a visually guided action task. In our monitoring task, the time interval between successive letter changes was randomly sampled from a uniform distribution ranging from 1.5 to 6.5 seconds (s). Thus, following a letter change, there was a 1.5 s 'silent period', during which the next letter change could not occur. Following this silent period, the probability of the next letter change (i.e., the hazard rate) linearly increased. With respect to our second hypothesis, we predicted that our participants would exploit this silent period, or more broadly, periods of low letter change probability, to transiently shift their gaze resources towards the manipulation task.

Unlike the visual monitoring of environmental events, where timing demands on central vision are typically externally determined, individuals would, in principle, be able to adjust the timing of their own actions and, consequently, the timing of the required action task fixations. With respect to our third hypothesis, we predicted that individuals would reduce competition for gaze resources between tasks by adjusting the timing of the action task. Importantly, this hypothesis assumes that participants would not only learn the statistical properties of letter changes in the monitoring task, but also possess knowledge of when and where action task fixations are required during the unfolding action task.

Results

Eleven participants performed, at their own pace, 30 consecutive trials of the ball-drop task in each of four experimental conditions. Participants performed the task using either their fingertips or tweezers, either as a standalone activity (referred to as 'single task' conditions) or concurrently with the visual monitoring task (referred to as 'dual task' condition). In each ball-drop trial, participants reached for and grasped a ball positioned on a platform adjacent to the base of a vertical tube. They then transported the ball to one of three slots within the tube, inserted it, and released it before returning their hand to its starting position. After the ball was released, it descended through the tube and returned to its starting position on the platform. One second after the ball returned to the start position (or was already located in the start position in the first trial), an auditory signal instructed the participant about which slot to use.

Figure 1A provides a view of the experimental setup from the participant's perspective. For a trial involving tweezers, it illustrates the path of the end-effector, which took place in a plane parallel to the participant's coronal plane, situated in front of the body. Each ball-drop trial was decomposed into seven consecutive action phases, distinguished by distinct kinematic events observed in the behaviour of the end-effector. These phases are defined in Figure 1B. The visual monitoring task involved detecting changes in a letter displayed on a text screen positioned in the upper right quadrant of the participant's scene (Fig. 1A; for details see Methods). The period between letter changes was randomly chosen from a uniform distribution, ranging from 1.5 to 6.5 s. Participants received rewards for successful ball drops and penalized for failing to detect letter changes, which were signaled by an auditory tone and visual feedback on the display.

We will first examine the coordination of gaze and end-effector movements in the single task conditions performed using either fingertips or tweezers. These conditions serve as baselines for comparison with the corresponding dual task conditions, which we will examine afterwards.

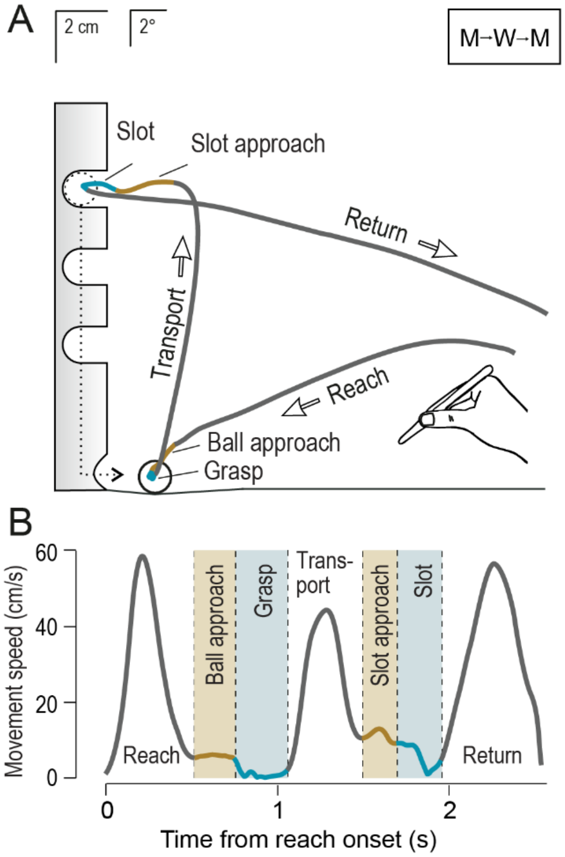

Figure 1. Apparatus and action phases in the ball-drop task. (A) Experimental setup from the participant's perspective. Illustrated is an example path of the end-effector tip during a tweezer trial, occurring in the work plane parallel to the coronal plane and situated 40 cm from the participant's eyes. (B) The corresponding velocity profile of the end-effector. (A, B) The task was segmented into 7 consecutive action phases separated by distinct kinematic events (see Methods). (1) *Reach phase*: starts when the hand leaves its starting position and is characterized by a bell-shaped velocity profile. (2) *Ball approach phase*: starts at a minimum (or inflexion point) in the velocity profile. (3) *Grasp phase*: starts at first contact with the ball. (4) *Transport phase*: starts when the ball is lifted from the platform and characterized by a bell-shaped velocity profile. (5) *Slot approach phase*: starts at a minimum (or inflexion point) in the speed profile and features low movement speed. (6) *Slot phase*: starts when the tips of the end-effector holding the ball enter the slot. (7) *Return phase*: starts when the ball is released and exhibits a bell-shaped velocity profile. The trial concludes when the hand returns to its original position.

Figures 2 A-B show gaze and end-effector paths for single trials performed with the fingertips and tweezers, respectively. With both end-effectors, participants typically fixated the ball as they reached toward it and, around the time the ball was grasped, shifted their gaze to the slot. Gaze remained at the slot until around the time the ball was dropped and then shifted back to the ball's start position at the base of the tube. Figures 2 C-D, which combine all trials from all participants, show, for each end-effector, the average speed of the tip of the end-effector (top) and the instantaneous probabilities of gaze fixating the ball at its start position (ball fixation) and the target slot (slot fixation), as a function of time. To temporally align trials while preserving information regarding action phases, we normalized the duration of each phase in each trial to the median duration of that phase computed across all trials within each condition.

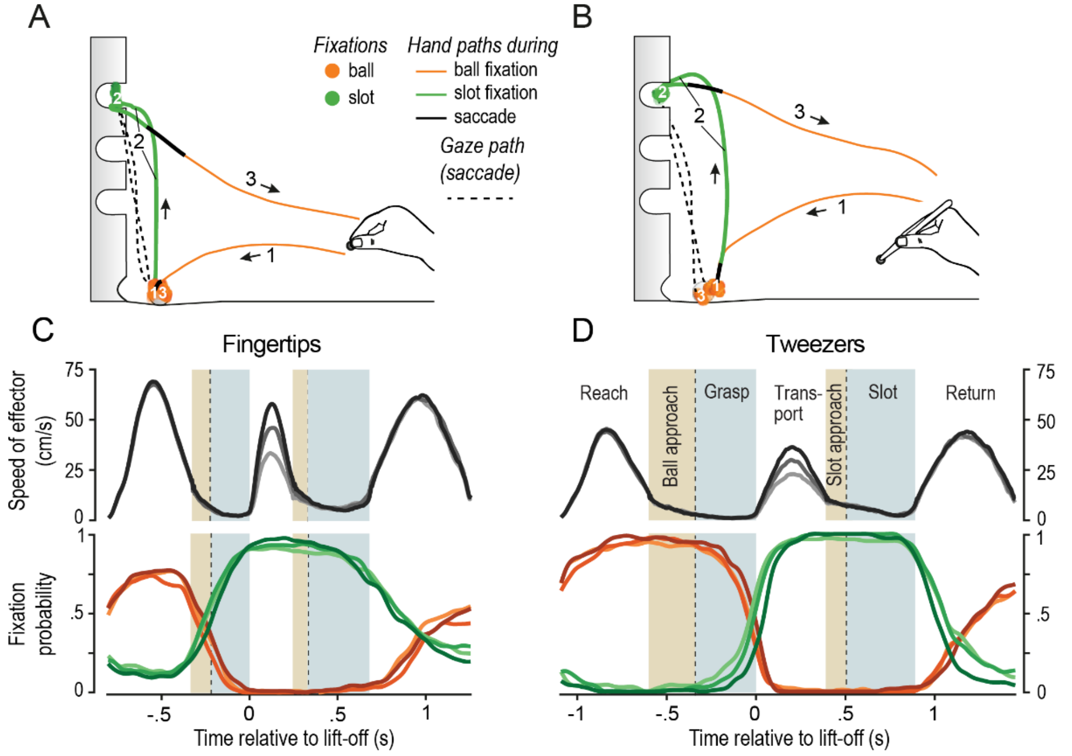

Figure 2. Gaze-hand coordination in the single-task condition. (A, B) Gaze and end-effector paths from exemplar trials performed with the fingertips (A) and tweezers (B). Fixations are color-coded by landmark (ball, slot), and end-effector paths (tip of the fingers or tweezers) are color-coded based on the current state of gaze (fixating the ball or slot or making a saccade). Numbers indicate the sequence of eye and hand movements. (C, D) Average speed of the end-effector (black) and the probabilities of fixating the ball (orange) and slot (green), shown as a function of time relative to lift-off, for fingertip (C) and tweezer (D) trials. Separate curves are shown for each slot (top, middle, and bottom coded dark to light). The alternating white, brown, and blue regions show the different movement phases labelled in D. The plots combine all trials from all participants and the duration of each phase in each trial was normalized to the median duration of that phase. Note that fixations were almost always directed to the landmarks, however, the sum of the probabilities of these fixations could be less than 1 due to saccades between the landmarks.

In both fingertip and tweezer trials, participants predominantly fixated the ball throughout the reach phase, although in fingertip trials gaze was sometimes directed towards the slot. Similarly, during most of the transport phase, participants predominantly fixated the slot. The timing of the gaze shift from the ball to the slot differed between fingertip and tweezer trials. In fingertip trials, this gaze shift occurred just before contact (-0.06 ± 0.05 s; mean ± sem), typically during either the late reach phase or the ball-approach phase. In contrast, in tweezer trials, this gaze shift occurred well after ball contact (0.29 ± 0.11 s), mainly during the grasp phase, and significantly later (t_{10} = 13.24; $p < 0.001$; d =

3.99) than in fingertip trials. This finding aligns with our prediction that establishing a stable grasp on the ball with tweezers required greater reliance on central vision compared to fingertips. In most fingertip and tweezer trials, gaze remained at the slot throughout the slot phase, before shifting to the ball start position. We observed that participants completed the ball-drop trials more rapidly (t_{10} = 4.94, $p < 0.001$; $d = 1.49$) when using fingertips (2.04 ± 0.33 s) compared to tweezers (2.53 ± 0.31 s). The greater time required to perform the task with tweezers resulted from increased durations of the ball approach, grasp, and transport phases ($p < 0.002$ in all three cases; separate paired t-tests for each action phase with p adjusted for multiple comparisons using the Holm-Bonferroni correction). These findings suggest that manipulating the ball posed greater challenges when using tweezers.

In the dual task conditions, participants distributed their gaze fixations between the text display and the action-related landmarks, consistent with the fact that detecting letter changes in the visual monitoring task required central vision (Fig. 3A-B). Consequently, the likelihood of fixating the ball and slot, at any given time, during the ball-drop task performance was diminished compared to the single task, irrespective of which end-effector was utilized (Fig. 3C–D). Moreover, the durations of occurring action landmark fixations were consistently shorter.

Propensity and patterns of action fixations depend on the end-effector used

When using the fingertips, participants primarily fixated the display during the reach, ball approach, and grasp phases (88% of trials). This indicates that grasping the ball could generally be accomplished without relying on central vision. Even in trials in which participants fixated the ball, these fixations occurred during the reach phase and gaze most often shifted away from the ball before the ball approach phase and almost never remained on the ball after contact. Transporting the ball and inserting it into the slot could also be accomplished while gaze remained on the display. However, participants briefly fixated the slot in about half of the trials (51%), as in the example shown in Fig. 3A. The probability of fixating the slot peaked midway through the transport phase, before steadily decreasing, and eventually approaching zero by the end of the slot phase (Fig. 3C). These findings suggest, in fingertip trials, inserting the ball into the slot could often be performed without the involvement of central vision. Nevertheless, the likelihood of fixating the slot during the slot approach and slot phase was higher than the likelihood of fixating the ball during the ball approach and grasp phase.

Because participants rarely fixated the ball and fixated the slot in approximate half of all trials, the two main gaze patterns observed in fingertip trials were ‘display-only’, where gaze remained on the display throughout the trial, and ‘slot’, where gaze shifted from the display to the slot and back to the display (see Figs. 3E-F). To examine the relationship between gaze pattern and manual performance, we compared the kinematic phase durations in display-only and slot trials using paired *t*-tests. We applied Holm-Bonferroni correction for multiple comparisons and only included participants (N=9) who demonstrated both fixation patterns. We found that the duration of the slot phase was shorter (t_8 = 3.76, adjusted p = 0.03) when participants fixated the slot (slot trials: 0.30 ± .054 s) compared to when they did not (display-only trials: 0.378 ± .056 s). No other phase duration was influenced by the gaze pattern (adjusted $p > 0.51$ in all cases).

In tweezer trials, participants were much more likely to fixate both the ball and the slot compared to fingertip trials (Fig. 3D). Participants almost always fixated the ball before ball contact (88% of trials), and the slot before slot entry (89% of trials). The probability of fixating the ball peaked towards the end of the reach phase and remained relatively high during the ball approach and most of the grasp phase. Similarly, the likelihood of fixating the slot peaked towards the end of the transport phase and remained fairly high during the slot approach phase and the slot phase. These findings suggest that, in tweezer trials, central vision was required in the vast majority of trials for both grasping the ball and inserting it into the slot.

In tweezer trials, the most prevalent gaze pattern was 'ball-display-slot', where participants shifted their gaze from the display to the ball, back to the display, then to the slot before returning to the display (as in the example shown in Fig. 3B). The second most common pattern was 'ball-slot', where participants shifted their gaze from the display to the ball and then directly to the slot before returning to the display (see Figs. 3E-F). To investigate the relationship between gaze pattern and manual performance, we compared the kinematic phase durations in the ball-display-slot and ball-slot trials using paired t-tests with a Holms-Bonferroni correction. Participants who had at least one trial of each fixation pattern (N = 8) were included in this analysis. We found that the transport phase was shorter (t_7 = 4.71, adjusted p = 0.01) when gaze shifted directly from the ball to the slot (ball-slot trials: 0.32 ± .051 s) compared to when gaze fixated the display between the ball and slot fixations (ball-display-slot trials: 0.551 ± .168 s). No other phase durations were affected by the gaze pattern (adjusted $p > 0.22$ in all cases).

As in the single task conditions, in the dual task conditions the ball-drop task was performed more slowly with the tweezers (2.77 ± 0.4 s) than with the fingertips (2.11 ± 0.23 s; t_{10} = 5.61, $p < 0.001$; d = 1.69). Paired t-tests with a Holms-Bonferroni correction revealed that, as in the single task, the greater time taken to perform the task with tweezers was due to increased durations of the reach, grasp, and transport phases (adjusted $p < .01$ in all three cases).

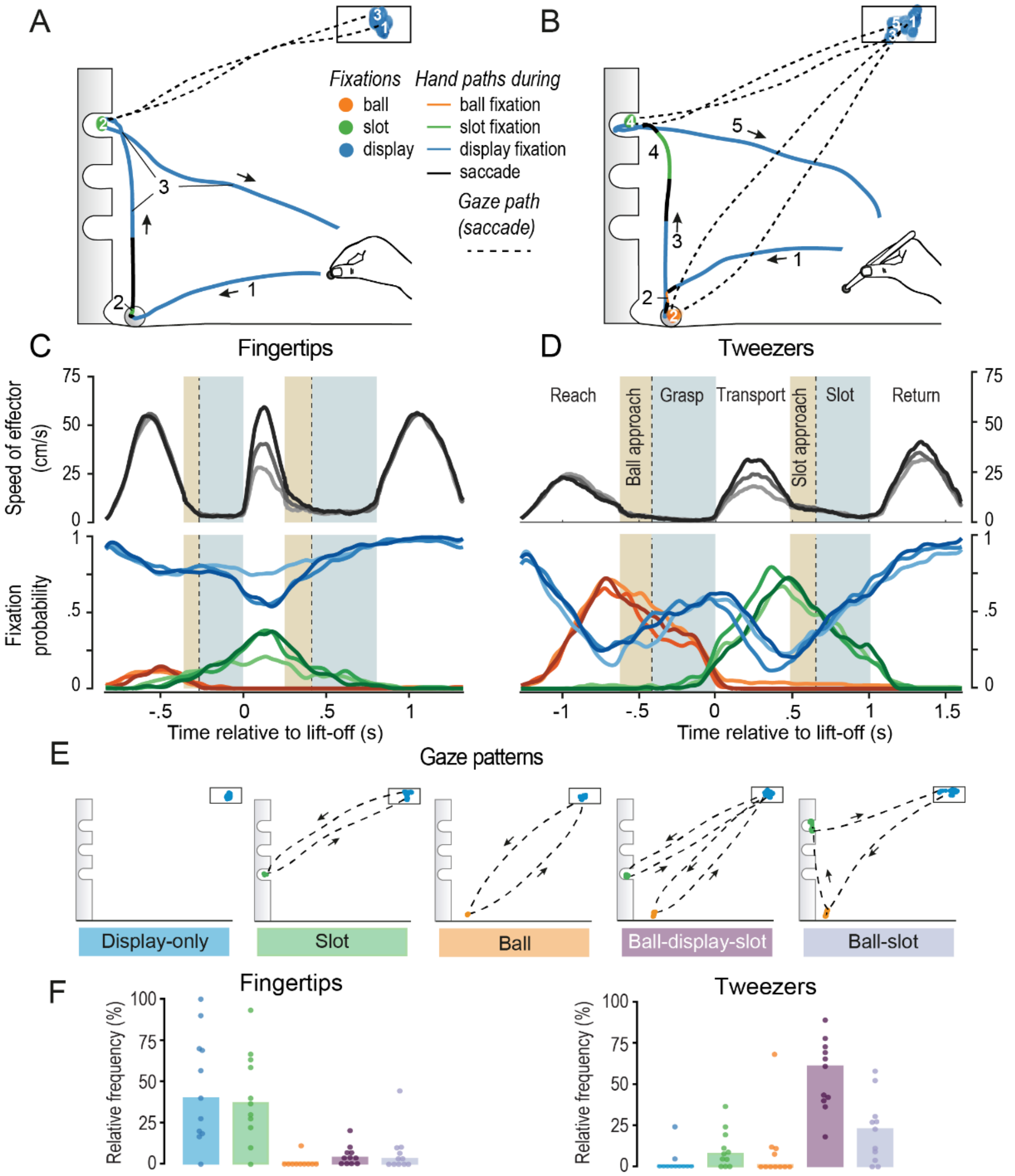

Figure 3. Gaze-hand coordination during dual-task conditions. (A, B) Gaze and end-effector paths from exemplar trials performed with the fingertips (A) and tweezers (B). Fixations are color-coded by landmark (ball, slot, display), and end-effector paths (tip of fingers or tweezers) are color-coded based on the current state of gaze (fixating the ball, slot or display, or making a saccade). Numbers indicate the sequence of eye and hand movements. (C, D) Average end-effector speed (black) and the probabilities of fixating the ball (orange), slot (green), and display (blue), shown as a function of time relative to lift-off, for fingertip (C) and tweezer (D) trials. Separate functions are shown for each slot (top to bottom coded dark to light). The alternating white, brown, and blue regions show the different movement phases labelled in D. The plots combine all trials from all participants and the duration of each phase in each trial was normalized to the median duration of that phase. (E) Five single-trial gaze patterns. (F) Mean percentage, averaged across participants, of each gaze pattern (color-coded as in E) in fingertip and tweezer trials. Dots represent individual participants.

Action fixations are anchored to contact events in the action task

The fixation probability functions shown above provide an overall view of the critical use of central vision in the ball-drop task (Fig. 3C-D). However, because these functions are based on time-normalized averaged data, they do not provide information on the trial-by-trial coordination between the timing of different action fixations and specific kinematic events. To examine this coordination, we carried out a linear regression analysis to determine whether the onset and offset times of ball and slot fixations could be predicted by the following kinematic events: (1) start of the reach phase, (2) start of the ball approach phase, (3) time of first ball contact (i.e., start of the ball grasp phase), (4) time of ball liftoff (i.e, start of the ball transport phase), (5) start of the slot approach phase and (6) time for slot entry (i.e., start of the slot phase). To reduce structural multicollinearity, these predictors were centered individually for each participant by subtracting the mean. Furthermore, we used study participants as a categorical nuisance factor to reduce variance related to the fact that participants performed the task at different speeds. Separate regression analyses were carried out for fixation onsets and offsets and for each action landmark (ball and slot) and end-effector (fingertips and tweezers). In all cases, the best predictor of fixation onset and fixation offset was the associated contact event.

In both fingertip and tweezer trials, we found that the initiation and termination of ball fixations were best predicted by the time of first ball contact, while the initiation and termination of slot fixations were best predicted by the time of slot entry. With respect to fixation onset times, our analysis showed that in fingertip trials, the onset of ball fixation was solely predicted by the time of first ball contact ($t_{1,32} = 2.84$; $p = 0.008$), while the onset of slot fixation was solely predicted by the time of slot entry ($t_{1,158} = 9.28$; $p < 0.001$). Similarly, in tweezer trials, the onset of ball fixation was primarily predicted by the time of first ball contact ($t_{1,243} = 6.26$; $p < 0.001$), and the onset of slot fixation was solely predicted by the time of slot entry ($t_{1,244} = 11.2$; $p < 0.001$). Comparable patterns were observed for the offset times of ball and slot fixations. In fingertip trials, the offset of ball fixation was solely predicted by the time of first ball contact ($t_{1,32} = 3.08$; $p = 0.004$), while the offset of slot fixation was best predicted by the time of slot entry ($t_{1,158} = 8.30$; $p < 0.001$). Similarly, in tweezer trials, the offset of ball fixation was best predicted by the first ball contact ($t_{1,243} = 5.18$; $p < 0.001$) while the offset of slot fixation was solely predicted by the time of slot entry ($t_{1,244} = 9.04$; $p < 0.001$). Thus, for both end-effectors, ball and slot fixations were closely coupled, in time, to ball and slot contact events, indicating a strong temporal linkage between the initiation and termination of action landmark fixations and their associated contact events.

Figure 4 supports the above regression analyses by showing that the ball and slot fixation periods, relative to ball contact and slot entry respectively, remained quite consistent across trials. The left panels of Figs. 4A and B show the timing of selected action phases and ball fixations relative to the time of first ball contact (time = 0). Each row represents a trial and all trials from all participants are shown. The orange lines depict periods of ball fixation. The onsets of the reach, ball approach and transport phases in each trial are marked by dots. Note that the trials are sorted by ball approach phase duration. Similarly, the left panels of Figs. 4C and D, show the timing of selected action phases and slot fixation relative to the time of slot entry. The green lines depict periods of slot fixation. The onsets of the transport, slot approach, and return phases in each trial are marked by dots. These trials are sorted by slot approach phase duration. In some trials, the ball or slot were re-fixated (see purple lines in Figs. 4A-D). This could occur when multiple attempts were needed to grasp and lift the ball or insert it into the slot. Note that the duration of both the ball and slot approach phases could vary considerably across trials.

The scatter plots on the right side of each panel in Fig. 4 show the position of each trial in its test block (No. 1 – 30), ranked by the duration of the ball or slot approach phase, as well as the position of each participant's trials in the same ranking (Pt. Nos. 1 -11). The scatter plots are marked by coloured (orange or green) and black dots, denoting trials with and without an action landmark fixation (ball or slot), respectively. The lack of apparent structure in these scatter plots suggests that neither the decision to fixate the action landmark nor the variation in ball and slot approach phase durations were influenced by trial position. Furthermore, these scatter plots suggest that, overall, participants exhibited similar behavior.

The timing of action fixation onsets, relative to contact events, showed remarkable consistency across action landmarks and end-effectors. Both ball and slot fixations typically began approximately 0.4 s before ball contact and slot entry, respectively (shown by the solid line curves in Figs. 4E and F). The timing of action fixation offsets was also consistent across action landmarks but influenced by the end effector used. In fingertip trials, gaze tended to shift away from both the ball and the slot before the contact event, with an average lead time of about 0.15 s (shown by dashed line curves in Fig. 4E). Conversely, in tweezer trials, the corresponding gaze shifts typically occurred shortly after the contact event, with an average lag of about 0.05 s (illustrated by dashed line curves in Fig. 4F).

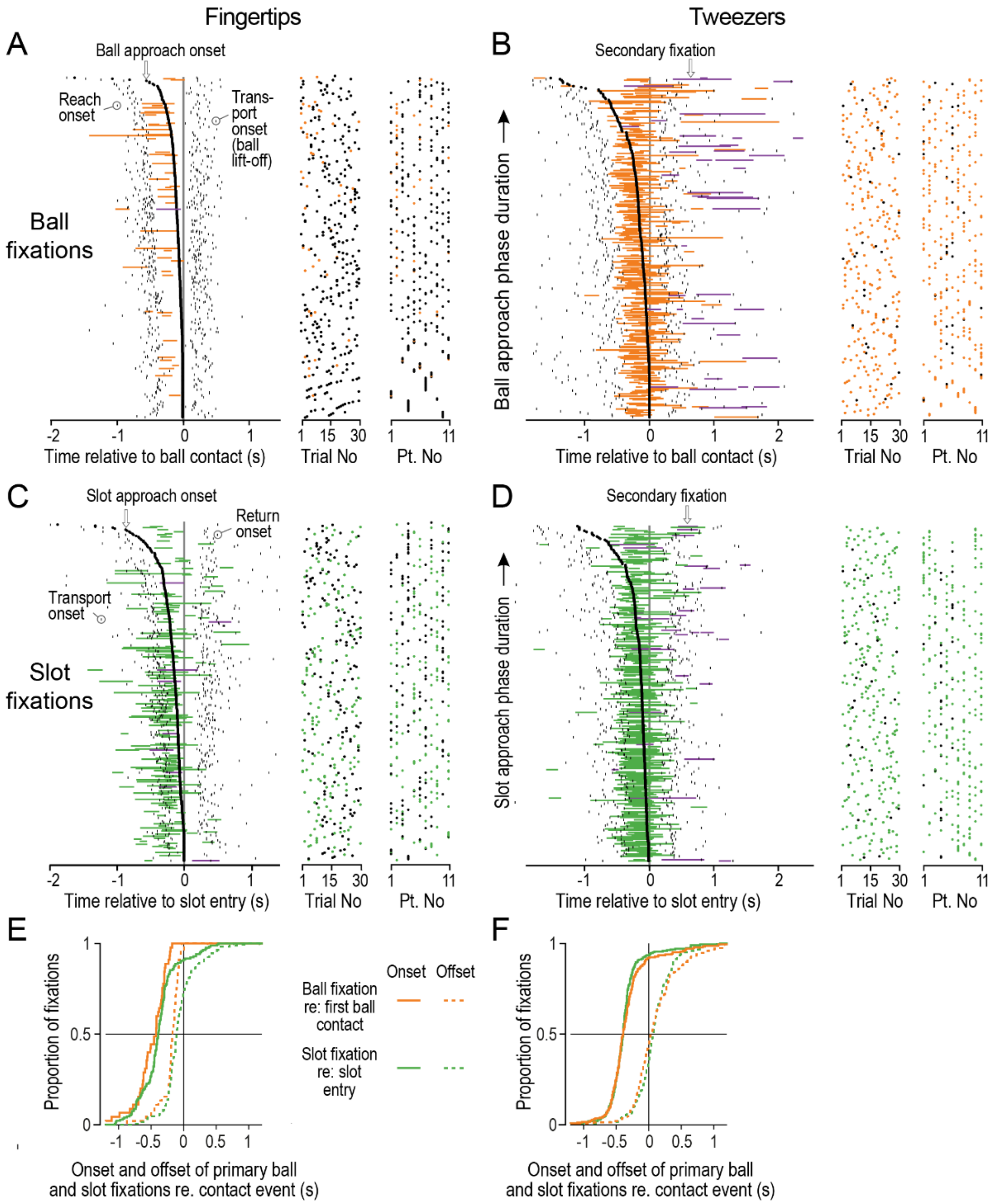

Figure 4. Timing and duration of action fixations relative to movement phases. The figure shows trials performed by all participants. (A, B) The left column shows time periods of initial (orange horizontal lines) and secondary (purple horizontal lines) ball fixations, aligned to the time of initial ball contact (gray vertical line), in fingertip (A) and tweezer (B) trials. The onsets of the reach and transport phases are marked by small black dots and the onset of the ball approach phase is marked by larger black dots. Data are sorted by the duration of the ball approach phase. The middle and right columns show each trial number and participant number, with orange and black dots depicting trials with and without a ball fixation. (C, D) Corresponding plots for initial (green) and secondary (purple) slot fixations, aligned to the time of slot contact. The small black dots mark the onsets of the transport and return phases and the larger black dots indicate the onset of the slot approach phase. Data are sorted by the slot approach phase duration. In the middle columns, the green and black dots depict trials with and without a ball fixation. (E) Cumulative distributions of ball and slot fixation onsets and offsets, aligned to the initial ball contact and slot entry respectively, in fingertip trials. (F) Corresponding distributions in tweezer trials.

The function of action landmark fixations can vary across trials

The variability in the timing of ball and slot fixations with respect to their related action phases suggests that their functions, in terms of *directing* and *guiding*, may differ across trials. We examined the function of each individual ball and slot fixation, recognizing that an individual fixation could serve multiple functions. A fixation was considered to be involved in *directing* if the ball or slot was fixated for at least 100 ms during the reach or transport phase, respectively. Similarly, a fixation was considered to be involved in *guiding* if the ball or slot was fixated for at least 100 ms between the start of the ball or slot approach phase and the end of the grasp or slot phase, respectively (i.e., the combined approach and manipulation phases). In addition to *directing* and *guiding*, gaze can also be engaged in '*checking*' the completion of action phases linked to a given landmark (Säfström et al., 2014). A fixation was considered to be involved in *checking* if the ball or slot was fixated for any period of time after the end of the grasp or slot phase, respectively. Figure 5A provides illustrative examples of slot fixations in tweezer trials demonstrating these different functions.

In fingertip trials, ball fixations were mainly involved in *directing* the end effector whereas slot fixations were also quite frequently involved in *guiding* (Fig. 5B). In contrast, in tweezer trials, both ball and slot fixations were approximately equally engaged in *directing* and *guiding* (Fig. 5C). Furthermore, a small proportion of fixations in both fingertip and tweezer trials were involved in *checking* (Fig. 5B and C). It is worth noting that in fingertip trials, most of the ball and slot fixations served only one function (80% overall), whereas in tweezer trials, this proportion was lower (48% overall) (see thin solid bars within each wide bar in Figs. 5B and C).

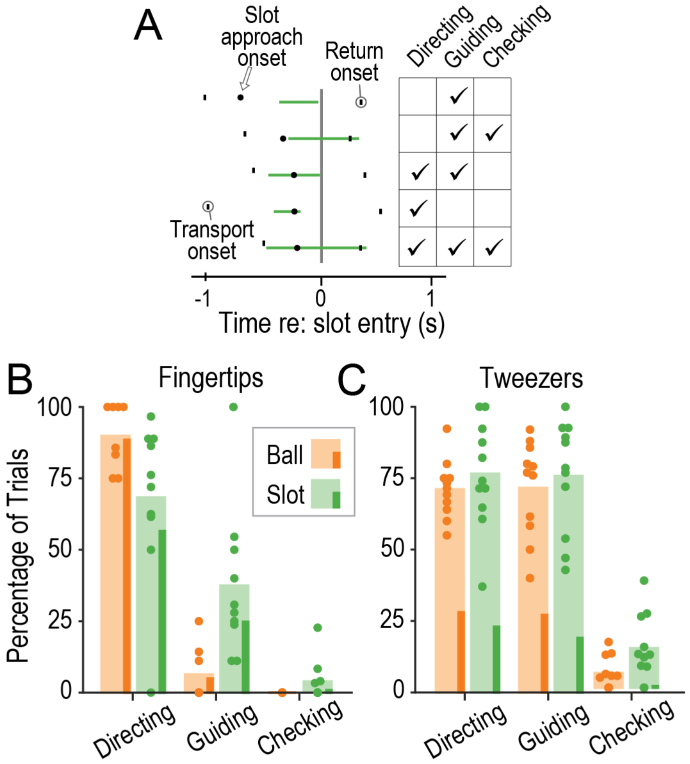

Figure 5. Classification of fixation functions. (A) Example of slot fixations from tweezer trials that serve different functions. (B, C) Wide bars represent the mean percentage, averaged across participants, of ball (orange) and slot (green) fixations in fingertip (B) and tweezer (C) trials engaged in directing, guiding, and checking. Note that a given fixation could be engaged in more than one function. Circles represent individual participants, and horizontal offsetting is used to show each participant (except for circles at zero). The thin bars represent the percentages of single-function fixations within each bar.

Monitoring task statistics influence task performance

In the following section, we investigate whether participants can learn and exploit the statistical properties of letter changes (LCs) to more efficiently distribute their gaze resources to the visual monitoring and manipulation tasks. We will demonstrate that participants adapt their gaze behavior both directly, by selecting different gaze patterns, and indirectly, through adjustments in manual behavior, based on LC statistics.

In the dual-task condition, the interval between LCs was drawn from a uniform distribution ranging from 1.5 to 6 s. This means that participants had a window of at least 1.5 s after detecting a LC to allocate gaze to the ball-drop task focus without the risk of missing the next LC. We will refer to this time window as the 'silent period'. Moreover, participants might also learn that the likelihood of the next LC gradually increases from 0 to 1 over the 5 s after the silent period, known as the hazard rate. Overall, participants performed well on the LC detection task, potentially allowing them to exploit these LC statistics. On average, there were 1.08 and 1.41 LCs per trial in fingertip and tweezer trials, respectively. Participants detected these LCs with 88.8 ± 11.8% and 87.1 ± 9.1% accuracy (mean ± standard deviation across participants).

To explore how LC statistics might affect gaze behavior, we studied whether the detection of LCs influenced the timing of ball and slot fixations. Specifically, we compared the frequency distributions of ball and slot fixation onsets—relative to the most recently detected LC before each fixation—with the expected distributions assuming fixation onsets occurred randomly with respect to LCs. We conducted separate analyses for trials associated with each of the main gaze patterns: ball-display-slot, ball-slot, and slot trials (Fig. 6A).

In both fingertip and tweezer trials in which participants fixated both the ball and slot (i.e., ball-slot and ball-display-slot trials), the distribution of ball fixations onsets deviated from the expected random distribution (Kolmogorov-Smirnov test, $p \leq 0.01$ in all four cases). Specifically, the frequency of ball fixation onsets during the silent period was notably higher than expected by chance (top row of Fig. 6A).

Notably, in tweezer trials, the choice of gaze pattern was strongly influenced by the timing of ball fixation onset relative to the preceding LC. When the ball was fixated during the silent period, participants were equally likely to use either the ball-slot or ball-display-slot pattern. However, if the ball was fixated after the silent period, the ball-display-slot pattern was almost always selected. That is, with few exceptions, participants only shifted their gaze directly from the ball to the slot (ball-slot pattern) if the ball was fixated within the silent period. Conversely, if the ball was fixated after the silent period, participants almost always fixated the display before fixating the slot (ball-display-slot pattern).

As expected, in both fingertip and tweezer trials in which both the ball and slot were fixated, the peak in the distribution of ball fixation onsets was followed by a subsequent peak in slot fixations distribution (bottom row of Fig. 6A). This occurred because in the great majority of these trials (96.4%), the last LC detected before the ball fixation was also the last LC before the slot fixation. In tweezer trials with the ball-display-slot gaze pattern, the additional peak in the distribution of slot fixation onsets during the silent period represents trials in which a LC was detected when gaze was at the display before shifting to the slot shortly afterwards. In contrast, in both fingertip and tweezer trials in which only the slot was fixated (slot-only gaze pattern), the distribution of slot fixation onsets did not differ significantly (KS test, $p > 0.06$ in both cases) from the expected random distribution (bottom row of Fig. 6A).

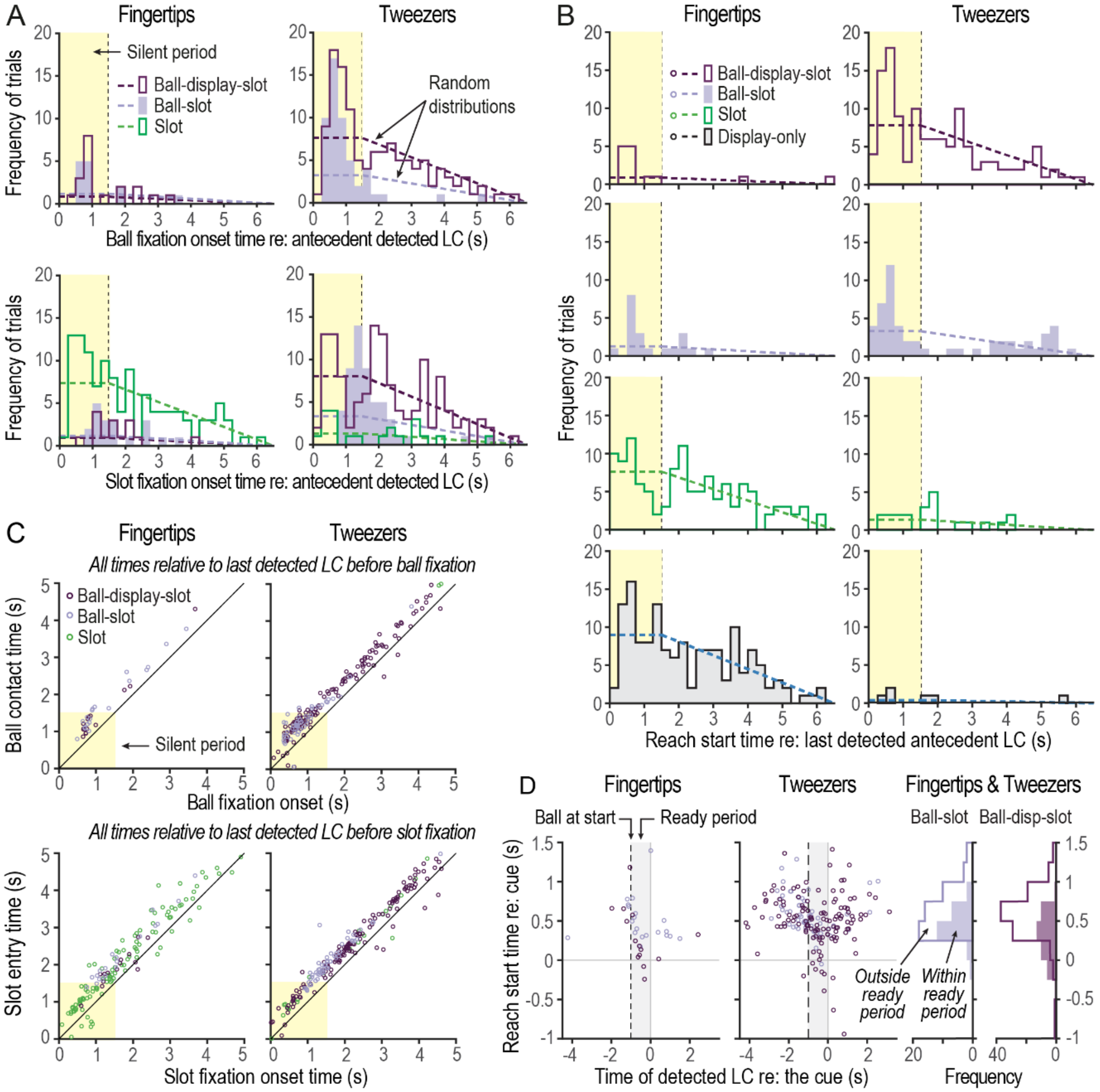

Figure 6. Relationship between eye and hand movements and letter changes (LCs). (A) Frequency distributions, combining all participants, of ball (top) and slot (bottom) fixation onsets—relative to the time of the last detected LC before the fixation onset—in fingertip (left) and tweezer (right) trials. The yellow region in each panel shows the silent period. Separate distributions shown for trials with the main gaze patterns. Dashed lines show the expected distributions assuming fixation onsets occurred randomly and thus independently of the timing of LCs. The expected frequency is constant within the silent period, during which the hazard rate (the probability of a LC occurring if one has not yet occurred) remains at 0, and then decreases, at a constant rate, over the next 5 s as the hazard rate increases from 0 to 1. (B) Corresponding frequency distributions of reach start times, relative to the last detected antecedent LC, in fingertip (left) and tweezer (tweezer) trials. Separate plots are shown for the four main gaze patterns. (C) Relationships between ball fixation and ball contact onset times (top), and between slot fixation and slot entry onset times (bottom) in fingertip (left) and tweezer (right) trials. Times relative to the last detected LC before fixation onset. Dots represent trials from all participants and are colour-coded by gaze pattern. (D) Relationship between reach start time and time of nearest detected LC, both relative to cue onset, in fingertip and tweezer trials with ball-slot and ball-display-slot patterns. Right panels show corresponding frequency distributions, combining fingertip and tweezer trials, of reach start times for trials with ball-slot and ball-display-slot trials. Separate distributions are shown for trials in which the LC was within or outside the 'ready period' (1 s period prior to the cue).

We also discovered that the LC statistics had an impact on manual performance in the ball-drop task, particularly concerning the timing of reach initiation (Fig. 6B). In both fingertip and tweezer trials in which participants fixated the ball (i.e., the ball-slot and ball-display-slot gaze patterns), the distribution of reach onset times, relative to the antecedent detected LC, differed (K-S test, $p < 0.05$ in both cases) from the expected distributions, assuming reach onset times occurred randomly with respect to LCs. Specifically, reach onsets were biased towards the silent period (top two rows of Fig. 6B), aligning with the bias observed in ball fixations. In contrast, in fingertip trials in which the ball was not fixated, including trials with the slot and display-only gaze patterns, the distribution of reach onset times did not differ from the expected random distribution (K-S test, $p \geq 0.4$; bottom two rows of Fig. 6B). We did not analyze the corresponding distributions for the slot and display-only gaze patterns in tweezer trials due to the limited number of observed trials.

These findings suggest that participants adopted a strategy to preferentially fixate the ball during the silent period by choosing to reach for the ball during this time frame. This strategy reduces the risk of failing to detect LCs while maintaining functional gaze-hand coordination. Indeed, we observed a close temporal relationship between ball fixation and ball contact as well as between slot fixation and slot entry across all gaze patterns involving action landmark fixations (Fig. 6C). The intercept and slope of the relationship between ball fixation onset and ball contact time were 0.434 s and 1.003 in fingertip trials and 0.411 s and 0.969 in tweezer trials, and the intercept and slope of the relationship between slot fixation onset and slot entry time were 0.368 s and 0.997 in fingertip trials and 0.380 s and 0.985 in tweezer trials. These intercepts align with our observation, noted above, that both ball and slot fixations began approximately 0.4 s prior to the contact event, irrespective of the end-effector used (Figs. 4E and F).

In the ball drop task, participants most often initiated their reach movement towards the ball after hearing the auditory cue that indicated the active slot in that trial. These ‘reactive reaches’ occurred about 0.5 s after the cue. However, in a substantial proportion of trials, participants initiated their reach in anticipation of the cue, such that the reach started either before the cue or shortly after the cue (and less than 0.5 s after the cue). If participants generated these ‘anticipatory reaches’ in response to a LC occurring shortly before the cue, and fixated the ball when doing so, it would explain the greater-than-expected frequency of both reach onsets and ball fixation onsets during the silent period in trials with the ball-slot and ball-display-slot gaze patterns.

To examine this further, we analyzed the relationship between the timing of reach onset, relative to the cue, and the timing of the detected LC, relative to the cue. We focused on fingertip and tweezer trials with the two most common gaze patterns involving ball fixation (ball-slot and ball-display-slot gaze). For each trial, we selected the detected LC closest in time to the midpoint of the 'ready period'—the 1-second interval between when the ball returned to its starting position and when the cue was given. Note that this LC could therefore occur either before or after the midpoint of the ready period. We observed that most anticipatory reaches, characterized by relatively small or negative reach onset times relative to the cue, occurred when the detected LC happened during the ready period (see the left two panels of Fig. 6D which shows reach start times, relative to the cue, plotted against the time of the detected LC, relative to the cue). This suggests that the decision to initiate reaching in anticipation of the cue is linked to the detection of a LC during the ready period.

The frequency distributions of reach start times, relative to the cue, in trials with the ball-slot and ball-display-slot patterns—depicted in the right two panels of Fig. 6D—revealed earlier reach start times when the LC occurred within the ready period, compared to when it occurred outside of it. Note that due to the relatively small number of fingertip trials, we combined fingertip and tweezer trials in these distributions. Importantly, for both gaze patterns, the distributions within and outside the ready period differed (KS test, $p < 0.02$ in both cases). This suggests a distinct influence of the timing of detected LCs on reach initiation during the ball drop task.

Overall, these results demonstrate two ways in which participants took advantage of the statistical properties of LCs to effectively reduce the competition between tasks for gaze resources. First, they modulated the timing of their reaching movements to preferentially fixate the ball during the silent period. Second, they selected gaze patterns, on a trial by trial basis, that increased the probability that gaze could be allocated to the action task with little or no cost in terms of the LC monitoring task.

Discussion

The broad aim of this study was to examine how people coordinate their eye and hand movements when performing a visually guided object manipulation task in parallel with a visual monitoring task that competes for central vision. Using this novel experimental approach, we tested three hypotheses related to how participants might optimize the allocation of gaze resources across tasks. These hypotheses concerned the timing and location of fixations directed to the action task, whether participants could learn and take advantage of the temporal regularities of the monitoring task when allocating gaze, and whether participants would modify the timing of their hand movements, based on these temporal regularities, to effectively reduce the competition between two tasks. We found support for all three hypotheses, which we will consider in turn.

Frequency of action task fixations

We found, as expected, that when the ball-drop task was performed in isolation, using either the fingertips or tweezers, gaze was directed exclusively to the ball and slot, with gaze arriving ahead of the hand or tool and departing around the time the hand or tool arrived or shortly afterward. This finding is consistent with previous research on eye-hand coordination in visually guided action tasks (Flanagan and Johansson, 2003; Fooken et al., 2021; Hayhoe, 2017; Johansson et al., 2001; Land et al., 1999). In contrast, we expected that when performing the ball-drop and monitoring tasks in parallel, gaze would be briefly allocated to the action task when visuomotor control is most critical. Consistent with this expectation, we found that the ball and slot were almost always fixated in tweezer trials but that fixations of the ball, especially, and slot were often not observed in fingertip trials. The increased use of central vision when controlling the tweezers was expected for several reasons. First, because the contact surfaces of the tweezer tips are smaller than the surfaces of the fingertips, greater spatial precision is required, particularly when grasping the ball. Second, the tweezer tips are more rigid than the fingertips, and therefore cannot mold around the ball, leading to a far less stable grasp. Soft contact surfaces are typically used in robotic manipulators to increase grasp stability and lower spatial precision requirements (Bicchi, 2000; Bicchi and Kumar, 2002; Billard and Kragic, 2019). Third, the tweezer tips offer limited tactile feedback regarding the contact state, and impaired tactile sensibility of the fingertips is known to increase reliance on visual feedback for object manipulation (Brink and Mackel, 1987; Chemnitz et al., 2013; Jenmalm and Johansson, 1997; Jerosch-Herold, 1993). Indeed, our results suggest that the analysis of gaze control when performing object manipulation tasks can provide a means of assessing tactile impairments, as well as the effectiveness of tools in terms of transmitting tactile information to the user, an important consideration in teleoperation tasks and robot assisted surgery.

Timing of action task fixations

Independent of the end-effector employed, we found that the timing of ball and slot fixation onsets and offsets were most closely correlated with ball contact and slot entry, respectively, in comparison to all other kinematic events. For both action task landmarks, and for both end-effectors, gaze arrived approximately 0.4 s prior to contact on average. However, consistent with our hypothesis that the functions served by action task fixations would differ with the end-effector employed, we found that gaze shifted away from the ball and slot well ahead of contact (0.15 s on average) in fingertip trials but just after contact (0.05 s on average) in tweezer trials. This timing is consistent with our findings that when using the fingertips, ball and slot fixations are primarily involved in *directing* the hand (or object in hand) to the landmark using peripheral vision, whereas when using tweezers, these fixations are also involved in *guiding* the hand with central vision as it approaches the landmark. Importantly, once the tips of the fingers or tweezers contact the ball or slot, tactile feedback becomes available, marking a transition between visuomotor and haptic sensorimotor control. This transition in the mode of sensorimotor control can be linked to the transition from motion control to force control that has been proposed to involve distinct control processes (Casadio et al., 2015; Chib et al., 2009; Kolesnikov et al., 2011; Piovesan et al., 2019). Note that in the ball drop task, tactile information is used not only to guide forces—as when grasping the ball—but can also drive kinematic adjustments—as when adjusting the position of the ball when inserting it into the slot. Importantly, in manipulation tasks, tactile information can be used to rapidly (90-120 ms) adjust both forces (Johansson and Flanagan, 2009) and kinematics (Pruszynski et al., 2018, 2016) through automatic feedback control processes.

Flexibility of gaze patterns

We found that across both fingertip and tweezer trials, participants used different gaze patterns when performing the ball drop task in parallel with the monitoring task. In fingertip trials, we observed two main gaze patterns—display only and slot only trials—distinguished by whether or not the participant opted to fixate the slot or keep gaze on the display throughout the trial. In tweezer trials, we also observed two main gaze patterns—ball-slot and ball-display-slot trials—distinguished by whether or not the participant decided to fixate the display between fixating the ball and slot. Importantly, in both fingertips and tweezer trials, the choice of gaze pattern was linked to task performance. In fingertip trials, the duration of the slot phase was shorter when participants fixated the slot, and in tweezer trials, the duration of transport phase was shorter when the gaze shifted directly from the ball to the slot, skipping the display. These results suggest that there is a trade-off between the action and visual monitoring tasks, where allocating gaze resources to the action task improves performance but comes at the risk of missing a letter change. However, as we will discuss next, participants can mitigate this trade-off by considering the LC statistics when deciding which gaze pattern—and associated kinematic performance—to select.

Modulation of gaze behaviour exploiting the statistics of the monitoring task

We hypothesized that participants would exploit the temporal statistics of events (i.e., LCs) in the visual monitoring task when making gaze allocation decisions. This hypothesis was supported in both fingertip and tweezer trials. During fingertip trials, the decision of whether or not to fixate the ball was strongly influenced by the LC statistics. Specifically, almost all of the ball fixations that were observed occurred during the silent period when the next LC could not occur. Similarly, during tweezer trials, the decision of whether or not to fixate the display—in between fixating the ball and slot—was influenced by LC statistics. Specifically, whereas participants often opted to skip the display when the ball fixation occurred within the silent period, they almost always fixated the display when the ball fixation occurred outside this period.

Our findings align with previous research on eye movements, illustrating that human gaze behavior is sensitive to probabilistic regularities in the environment (Jovancevic-Misic and Hayhoe, 2009). For example, individuals adjust the timing of their gaze shifts based on the learning of temporal statistics of relevant visual events to optimize event detection in two separate spatial locations where event durations vary independently (Hoppe and Rothkopf, 2016). In addition, during visual search tasks, individuals strategically allocate gaze based on the spatial statistics of their surroundings to efficiently explore (Eckstein, 2017; Hoppe and Rothkopf, 2019; Najemnik and Geisler, 2005; Renninger et al., 2007). Our study adds a distinct perspective by demonstrating that humans can learn and exploit the temporal patterns of externally determined events in the visual environment while concurrently engaged in an action task that relies on visual guidance. This suggests that the processes involved in using visual information in sensorimotor control—including peripheral and central vision and gaze-related signals—are largely independent of the processes involved in extracting statistical regularities from the visual environment.

Modulation of manual behaviour exploiting the statistics of the monitoring task

Unlike the visual monitoring of environmental events, where timing demands on central vision are typically externally determined, individuals would, in principle, be able to adjust the timing of their own actions. We hypothesized that participants would tune the timing of their manual actions to decrease competition for gaze resources between the ball drop and LC detection tasks. In support of this hypothesis, we observed that our participants adjusted the onset time of their reaching movements such that ball fixations—supporting the reaching movement and ball grasp—occurred during the silent period far more often than would be expected if reach timing was uncoupled from the LC statistics. Importantly, this result suggests that participants not only learned the statistical properties of letter changes in the monitoring task, but also possess knowledge of when and where action task fixations are required during the unfolding action task.

Interestingly, we found that participants tended to adjust the timing of their reaching movements (and thereby lower the probability of a LC occurring around the time of ball grasp) even when using their fingertips, despite the fact that the ball was seldom fixated. A possible interpretation of this finding is that visual attentional mechanisms used to monitor LCs interfere with 'visuomotor' attentional mechanisms used to direct the hand to targets in peripheral vision. Although the use of peripheral vision and gaze-related signals to *direct* the hand is most effective when foveating that reach target, these signals can also be used to *direct* the hand when foveating a location separate from the reach target (de Brouwer et al., 2018; Neggers and Bekkering, 2001, 2000).

Conclusion

The current paper provides novel insights into how eye and hand movements are controlled and coordinated in real-world action tasks. First, our results provide support for the hypothesis that, under conditions in which there is competition for gaze, participants prioritize key functions linked to control points—involving contact events between the hand, or tool in hand, and objects in the environment—when allocating gaze to action tasks. Second, our results support the hypothesis that participants learn the temporal regularities of the external environment and exploit this knowledge to improve task performance by adapting both their hand and eye movements.

Methods

Participants

Eleven right-handed students (8 male; aged 22 to 33 yr) participated in the study. All participants reported normal or corrected-to-normal vision and were naive to the purpose of the study. The study was approved by the ethics committee of the University of Umeå and participants gave written informed consent before participating in the study.

Apparatus and general procedure

Participants sat at a table on which the ball-drop apparatus was installed. The apparatus consisted of a 15 cm high vertically oriented Perspex tube (inner diameter = 14 mm; wall thickness = 3 mm) that was attached to the middle of a wooden platform. The Perspex tube was fixed about 2.5 cm to the left of the participant's mid-sagittal plane and the top of the tube was at participants' eye level. The tube had three slots centered around 5, 8, and 11 cm above the platform surface. The manual task was to reach for and grasp a small ball (12 mm diameter polished brass sphere) located on the platform, transport it into a prescribed target slot, drop it through the tube, and return the hand to its support (a horizontal plate located adjacent to the platform, extending 20 cm from its right end). The start position of the ball was located 3 cm to the right of the vertical midline of the tube and the movements took place in a frontal work plane at 40 cm distance. The platform surface was slanted (~1° slope) such that the ball rolled to its start position when exiting the tube. In different blocks of trials, the ball was grasped either with the fingertips or with a pair of tweezers held by the right hand as a pen. The tweezers, made of plastic, were 14 cm long and had cylindrical tips of 4 mm in diameter, coated for 12 mm with thin plastic tubing to increase the friction against the ball.

Task design

A trial began with a verbal, pre-recorded command (“bottom”, “middle”, or “top”) that instructed the participant into which slot to drop the ball and ended when the hand returned to its support. A new verbal command started the next trial 1 second following the instance the ball had rolled back to its start position after being dropped through the tube. The participants performed the task at a preferred speed. In the single task conditions, the participants performed only the ball-drop task (Fig. 1 A and B). In the dual task conditions, the participants performed a visual detection task that engaged foveal vision while concurrently performing the ball-drop task. The task was to detect a letter change (LC) on a LED text display located in the upper right quadrant of the scene (Fig. 1 C and D). After randomly distributed times, ranging between 1.5 and 6.5 s (uniform distribution), the letter M was changed to W for 300 ms and then back to M. The participants were instructed to report each M → W → M sequence by immediately pressing the button-switch held in the left hand. If the button was not pressed within 1 second after a change, it was considered as a miss. A brief computer-generated beep sound and flashing of hash marks on the display for 600 ms signaled to the participants that they had missed a sequence. The visual angle between the center of the letter area and the center of the top slot was 24°. The visual angle to the ball was 28°. The size of the displayed letters (M, W) corresponded to 0.5 ° × 0.7 ° visual angle.

To ensure that the participants relied on foveal vision to detect LCs rather than perceiving them as peripheral visual events, the letter M alternated its horizontal position by 0.6° visual angle at randomly distributed times, ranging between 1 and 3 seconds (uniform distribution). Pilot tests showed that subjects had to foveate the display to detect the occurrence of M → W → M changes. To motivate the participants to simultaneously perform both tasks, they received one Swedish krona (SEK) for each ball-drop and lost 3 SEK for each undetected LC. Next to the letter, the display showed continuously the monetary balance, which could not go below zero. Participants were informed before the tests about the gain and loss rules.

Task order

Participants performed four conditions in the following order: single-task with fingertips, single-task with tweezers, dual-task with fingertips, and dual-task with tweezers. In each condition, the participants performed ten trials directed to each of three slots resulting in 30 trials per condition. The target slot varied in an order unpredictable for the participants.

Data collection

Gaze position was recorded at 120 samples/s using an infrared video-based eye-tracking system (RK-726PCI pupil/corneal tracking system, ISCAN Inc., Burlington, MA). An adjustable chin support stabilized the head together with a forehead support to which the head was strapped by Velcro tape. The standard deviations of the error distributions of gaze position measurements in the horizontal and vertical direction were 0.50° and 0.52° of visual angle (or 0.35 and 0.36 cm in the work plane), respectively. Miniature electromagnetic position-angle sensors with six degrees of freedom (RX1-D miniature receiver; FASTRAK, Polhemus, Colchester, VT) recorded at 60 samples/s the position of the tip of the participant's right index finger and the tips of the tweezers. The fingertip sensor was attached to the nail and the position of the fingertip was represented as the site of preferred contact with the ball. That is, in calibration trials performed before the actual ball-drop trials, we offset electronically the sensor for the preferred contact site obtained when participants were asked to grasp the ball when located at its start position. The sensor of the tweezers, attached to their proximal end, was electronically offset to record the midpoint between their tips.

Signals from a six-axis force-torque transducer (Nano F/T transducer, ATI Industrial Automation, Apex, NC; sampling rate 400 Hz) was used to detect the first contact with the ball when reached for. The sensor was attached underneath a rectangular plate (14 x 45 mm) that was a part of the platform surface and extended laterally from the bottom of the tube to 9 mm beyond the start position of the ball. Signals from this sensor could also be used to detect when the ball was lifted off the platform and impacted on the platform after being dropped through the tube. An optical reflex detector (SG-2BC, Kodenshi, Japan) mounted in the hole at the ball start position indicated (digital signal) when the ball was at this location within 1 mm. Located at the lower edge of each slot, the same type of reflex detectors provided a digital signal when the ball had dropped about 5 mm. To estimate the position of the grasped ball when transported, we used the sensors that recorded the fingertip position and the tips of the tweezers. All data were sampled using the SC/ZOOM software (Physiology Section, IMB, Umeå University). The signals from the various sensors were time synchronized and stored at 200 samples/s using linear interpolation between consecutive samples.

Gaze analysis

We identified the position of gaze fixations in the work plane using previously described criteria (Johansson et al., 2001). To assess locations and timing of fixations we defined three critical fixation zones (centroid with a radius of 2.5 cm) around the ball, the selected slot, and the text display. Gaze had to be within a given fixation zone for at least 20 samples (100 ms) to be classified as a fixation. Unless indicated otherwise, fixation probability and timing were collapsed across slots. To assess sequences of eye movements throughout the trial, we defined five different gaze patterns: (1) 'display-only' where gaze remained on the display throughout the trial, (2) 'ball' where gaze shifted from the display to the ball and back to the display, (3) 'slot' where gaze shifted from the display to the slot and back to the display, (4) 'ball-slot' where gaze shifted from the display to the ball and then to the slot before returning to the display, and (5) 'ball-display-slot' where gaze shifted from the display to the ball, back to the display, and then to then slot before returning to the display. Trials in which there were multiple fixations of a landmark in a given trial were not classified.

Movement analysis

To describe the movement sequence in the ball-drop task we defined seven kinematic phases depending on the speed of the fingertips and tweezers. Speed was computed as the vector sum of the first time derivative of filtered horizontal and vertical position signals (2nd order Butterworth low-pass filter with a cut-off frequency of 10 Hz). Reaching for the ball and ball transport typically showed a primary large movement component with a nearly symmetric bell-shaped velocity profile, followed by more irregular movement components with lower peak velocities. For both the reach and the transport we used the first and second time differential of the movement speed to detect the notch in the speed signal that demarcated the instance of transition from the primary movement to the subsequent submovements. The onset of the *reach* phase was defined at the times at which the speed of the speed of the endpoint of the effector exceeded 2 cm/s. The offset of the reach and onset of the *ball approach* phase was defined by the notch in the speed profile of the fingertips or tweezers that followed the large, initial reach movement. The onset and offset of *ball grasp* were defined by the times at which the ball was contacted and lifted off the surface, respectively. Ball grasp was followed by the *transport* phase. The offset of the transport phase marked the onset of the *slot approach* phase and was defined by the notch in the speed profile of the fingertips or tweezers that followed the large transport movement. The onset of *slot entry* was defined as the instance the ball was 1 cm to the right of the position where it was released inside the tube and the offset of slot entry was defined as the time the ball was dropped as detected by an optical sensor in the tube. Finally, the slot entry phase was followed by the *return* phase and the offset of the return phase were defined at the times at which the speed of the effector dropped below 2 cm/s.

To calculate the speed of the effector, the fixation probability at a given landmark, the LC probability, and the probability of being in the silent period in a normalized time frame, we calculated the median duration of each movement phase. In each trial, end-effector velocity or probabilities were up- or down-sampled to match the median duration of each movement phase. We then calculated normalized movement speed or probabilities for each participant and slot. Note that we filtered probability traces of individual participants with a 2nd order Butterworth low-pass filter (cut-off frequency of 10 Hz) before averaging across participants. When generating these plots, the duration of the ball approach and ball grasp phases and slot approach and slot entry phases were combined because the duration of the approach phases were often very short (single samples).

Data exclusion

We excluded trials, in which participants dropped the ball after initiating the transport phase. Overall, we excluded 64 trials (4.7 %) and no more than 7 trials were excluded for any participants.

Statistical analyses

We assessed the effect of end-effector (fingertips vs. tweezers) and task condition (single vs. dual) using a repeated-measures ANOVA. Fixation locations and timing were directly compared between effectors and task conditions using Welch's two-sample paired *t*-tests. Distributions of actual and expected LCs were compared using a two-sample Kolmogorov-Smirnov (KS) test. Distributions of fixation pattern relative to LCs were compared to a uniform distribution using a one-sample KS test. The effect of fixation pattern of kinematic phase duration was tested using multivariate analysis of variance (MANOVA). To identify to which action phase ball and slot fixations were temporally coupled we ran a general linear model (GLM) with the onset of movement phases as fixed effects and participants as random effect:

$$t_{fixation\ time} \sim t_{reach} + t_{ball\ approach} + t_{grasp} + t_{transport} + t_{slot\ approach} + t_{slot\ entry} + (1|participant)$$

To reduce structural multicollinearity among the predictors, the onset of each movement phase was centered individually for each participant by subtracting the mean. All statistical analyses were conducted in R (R Core Team, 2022; www.r-project.org).

Citation diversity statement

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field (Bertolero et al., 2020; Caplar et al., 2017; Chatterjee & Werner, 2021; Dion et al., 2018; Dworkin et al., 2020; Fulvio et al., 2021; Maliniak et al., 2013; Mitchell et al., 2013; Wang et al., 2021; Zurn et al., 2020). Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman (Dworkin et al., 2020; Zhou et al., 2020). By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 29.41% woman(first)/woman(last), 11.76% man/woman, 8.82% woman/man, and 50.0% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained the predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of colour (Ambekar et al., 2009; Chintalapati et al., 2023). By this measure (and excluding self-citations), our references contain 4.49% author of colour (first)/author of colour(last), 12.87% white author/author of colour, 15.92% author of colour/white author, and 66.72% white author/white author. This method is limited in that a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

References

Ambekar A, Ward C, Mohammed J, Male S, Skiena S. 2009. Name-ethnicity classification from open sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49–58. Doi: https://doi.org/10.1145/1557019.1557032

Ballard DH, Hayhoe MM, Li F, Whitehead SD. 1992. Hand-eye coordination during sequential tasks. *Philos Trans R Soc Lond B Biol Sci* **337**:331–9. Doi: https://doi.org/10.1098/rstb.1992.0111

Bertolero MA, Dworkin JD, David SU, Lloreda CL, Srivastava P, Stiso J, Zhou D, Dzirasa K, Fair DA, Kaczkurkin AN, Marlin BJ, Shohamy D, Uddin LQ, Zurn P, Bassett DS. 2020. *Racial and ethnic imbalance in neuroscience reference lists and intersections with gender*. Doi: https://doi.org/10.1101/2020.10.12.336230

Bicchi A. 2000. Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity. *IEEE Trans Rob Autom* **16**:652–662. Doi: https://doi.org/10.1109/70.897777

Bicchi A, Kumar V. 2002. Robotic grasping and contact: A review. *In Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation*. Symposia Proceedings (Cat. No.00CH37065). Doi: https://doi.org/10.1109/robot.2000.844081

Billard A, Kragic D. 2019. Trends and challenges in robot manipulation. *Science* **364**. Doi: https://doi.org/10.1126/science.aat8414

Brink EE, Mackel R. 1987. Sensorimotor performance of the hand during peripheral nerve regeneration. *J Neurol Sci* **77**:249–266. Doi: https://doi.org/10.1016/0022-510X(87)90127-4

Caplar N, Tacchella S, Birrer S. 2017. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, **1**:0141. Doi: https://doi.org/10.1038/s41550-017-0141

Casadio M, Pressman A, Mussa-Ivaldi FA. 2015. Learning to push and learning to move: the adaptive control of contact forces. *Front Comput Neurosci* **9**:118. Doi: https://doi.org/10.3389/fncom.2015.00118

Chatterjee, P., & Werner, R. M. (2021). Gender Disparity in Citations in High-Impact Journal Articles. *JAMA Network Open*, **4**:e2114509. Doi: https://doi.org/10.1001/jamanetworkopen.2021.14509

Chemnitz A, Dahlin LB, Carlsson IK. 2013. Consequences and adaptation in daily life – patients' experiences three decades after a nerve injury sustained in adolescence. *BMC Musculoskeletal Disorders*. Doi: https://doi.org/10.1186/1471-2474-14-252

Chib VS, Krutky MA, Lynch KM, Mussa-Ivaldi FA. 2009. The separate neural control of hand movements and contact forces. *J Neurosci* **29**:3939–3947. Doi: https://doi.org/10.1523/JNEUROSCI.5856-08.2009

Chintalapati R, Laohaprapanon S, Sood G. 2023. Predicting Race and Ethnicity from the Sequence of Characters in a Name (No. arXiv:1805.02109). arXiv. http://arxiv.org/abs/1805.02109

de Brouwer AJ, Gallivan JP, Flanagan JR. 2018. Visuomotor feedback gains are modulated by gaze position. *J Neurophysiol* **120**:2522–2531.

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered Citation Patterns across Political Science and Social Science Methodology Fields. *Political Anal*, **26**:312–327. Doi: https://doi.org/10.1017/pan.2018.12

Dworkin JD, Linn KA, Teich EG, Zurn P, Shinohara RT, Bassett DS. 2020. The extent and drivers of gender imbalance in neuroscience reference lists. *Nat Neurosci* **23**:918–926. Doi: https://doi.org/10.1038/s41593-020-0658-y

Eckstein MP. 2017. Probabilistic Computations for Attention, Eye Movements, and Search. *Annu Rev Vis Sci* **3**:319–342. Doi: https://doi.org/10.1146/annurev-vision-102016-061220

Flanagan JR, Bowman MC, Johansson RS. 2006. Control strategies in object manipulation tasks. *Curr Opin Neurobiol* **16**:650–659. Doi: https://doi.org/10.1016/j.conb.2006.10.005

Flanagan JR, Johansson RS. 2003. Action plans used in action observation. *Nature* **424**:769–771. Doi: https://doi.org/10.1038/nature01861

Fooken J, Baltaretu BR, Barany DA, Diaz G, Semrau JA, Singh T, Crawford JD. 2023. Perceptual-Cognitive Integration for Goal-Directed Action in Naturalistic Environments. *J Neurosci* **43**:7511–7522. Doi: https://doi.org/10.1523/JNEUROSCI.1373-23.2023

Fooken J, Kreyenmeier P, Spering M. 2021. The role of eye movements in manual interception: A mini-review. *Vision Res*, **183**:81–90. Doi: https://doi.org/10.1016/j.visres.2021.02.007

Fulvio JM, Akinnola I, Postle BR. 2021. Gender (Im)balance in Citation Practices in Cognitive Neuroscience. *J Cog Neurosci*, **33:**3–7. Doi: https://doi.org/10.1162/jocn_a_01643

Goodale MA, Pelisson D, Prablanc C. 1986. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature* **320**:748–750. Doi: https://doi.org/10.1038/320748a0

Hayhoe MM. 2017. Vision and Action. *Annu Rev Vis Sci* **3**:389–413. Doi: https://doi.org/10.1146/annurev-vision-102016-061437

Hoppe D, Rothkopf CA. 2019. Multi-step planning of eye movements in visual search. *Sci Rep* **9**:144. Doi: https://doi.org/10.1038/s41598-018-37536-0

Hoppe D, Rothkopf CA. 2016. Learning rational temporal eye movement strategies. *Proc Natl Acad Sci* **113**:8332–8337. Doi: https://doi.org/10.1073/pnas.1601305113

Illamperuma NH, Fooken J. 2024. Towards a functional understanding of gaze in goal-directed action. *J Neurophysiol* **132**:767–769. Doi: https://doi.org/10.1152/jn.00342.2024

Jenmalm P, Johansson RS. 1997. Visual and somatosensory information about object shape control manipulative fingertip forces. *J Neurosci* **17**:4486–4499. Doi: https://doi.org/10.1523/JNEUROSCI.17-11-04486.1997

Jerosch-Herold C. 1993. Measuring Outcome in Median Nerve Injuries. *Journal of Hand Surgery*. Doi: https://doi.org/10.1016/0266-7681(93)90019-c

Johansson RS, Flanagan JR. 2009. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nat Rev Neurosci* **10**:345–359. Doi: https://doi.org/10.1038/nrn2621

Johansson RS, Westling G, Bäckström A, Flanagan JR. 2001. Eye-hand coordination in object manipulation. *J Neurosci* **21**:6917–6932. Doi: https://doi.org/10.1523/JNEUROSCI.21-17-06917.2001

Jovancevic-Misic J, Hayhoe M. 2009. Adaptive gaze control in natural environments. *J Neurosci* **29**:6234–6238. Doi: https://doi.org/10.1523/JNEUROSCI.5570-08.2009

Kolesnikov M, Piovesan D, Lynch KM, Mussa-Ivaldi FA. 2011. On force regulation strategies in predictable environments. *Conf Proc IEEE Eng Med Biol Soc* **2011**:4076–4081. Doi: https://doi.org/10.1109/IEMBS.2011.6091013

Kowler E. 2011. Eye movements: the past 25 years. *Vision Res* **51**:1457–1483. Doi: https://doi.org/10.1016/j.visres.2010.12.014

Land MF. 2006. Eye movements and the control of actions in everyday life. *Prog Retin Eye Res* **25**:296–324. Doi: https://doi.org/10.1016/j.preteyeres.2006.01.002

Land MF, Furneaux S. 1997. The knowledge base of the oculomotor system. *Philos Trans R Soc Lond B Biol Sci* **352**:1231–1239. Doi: https://doi.org/ 10.1098/rstb.1997.0105

Land M, Mennie N, Rusted J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* **28**:1311–1328. Doi: https://doi.org/10.1068/p2935

Maliniak D, Powers R, Walter BF. 2013. The Gender Citation Gap in International Relations. *International Organization*, **67**:889–922. Doi: https://doi.org/10.1017/S0020818313000209

Mitchell SM, Lange S, Brus H. 2013. Gendered Citation Patterns in International Relations Journals. *Int Stud Perspect*, **14**:485–492. Doi: https://doi.org/10.1111/insp.12026

Najemnik J, Geisler WS. 2005. Optimal eye movement strategies in visual search. *Nature* **434**:387–391. Doi: https://doi.org/10.1038/nature03390

Neggers SFW, Bekkering H. 2001. Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal. *J Neurophysiol* **86**:961–970. Doi: https://doi.org/10.1152/jn.2001.86.2.961

Neggers SFW, Bekkering H. 2000. Ocular Gaze is Anchored to the Target of an Ongoing Pointing Movement. *J Neurophysiol*, **83**:639-651. Doi: https://doi.org/10.1152/jn.2000.83.2.639

Piovesan D, Kolesnikov M, Lynch K, Mussa-Ivaldi FA. 2019. The Concurrent Control of Motion and Contact Force in the Presence of Predictable Disturbances. *J Mech Robot* **11**:060903. Doi: https://doi.org/10.1115/1.4044599

Pruszynski JA, Flanagan JR, Johansson RS. 2018. Fast and accurate edge orientation processing during object manipulation. *eLife* **7**. Doi: https://doi.org/10.7554/eLife.31200

Pruszynski JA, Johansson RS, Flanagan JR. 2016. A Rapid Tactile-Motor Reflex Automatically Guides Reaching toward Handheld Objects. *Curr Biol* **26**:788–792. Doi: https://doi.org/10.1016/j.cub.2016.01.027

Renninger LW, Verghese P, Coughlan J. 2007. Where to look next? Eye movements reduce local uncertainty. *J Vis* **7**:6. Doi: https://doi.org/ 10.1167/7.3.6

Säfström D, Johansson RS, Flanagan JR. 2014. Gaze behavior when learning to link sequential action phases in a manual task. *J Vis* **14**. Doi: https://doi.org/10.1167/14.4.3

Saunders JA, Knill DC. 2004. Visual feedback control of hand movements. *J Neurosci* **24**:3223–3234. Doi: https://doi.org/10.1523/JNEUROSCI.4319-03.2004

Saunders JA, Knill DC. 2003. Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp Brain Res* **152**:341–352. Doi: https://doi.org/10.1007/s00221-003-1525-2

Sood G, Laohaprapanon S. 2018. Predicting race and ethnicity from the sequence of characters in a name. Doi:https://doi.org/10.48550/ARXIV.1805.02109

Wang X, Dworkin JD, Zhou D, Stiso J, Falk EB, Bassett DS, Zurn P, Lydon-Staley DM. 2021. Gendered citation practices in the field of communication. *Ann Int Commun Assoc*, **45**:134–153. Doi: https://doi.org/10.1080/23808985.2021.1960180

Yarbus AL. 1967. Eye Movements and Vision. Doi: https://doi.org/10.1007/978-1-4899-5379-7

Zhou D, Cornblath EJ, Stiso J, Teich EG, Dworkin JD, Blevins AS, Bassett DS. 2020. *Gender diversity statement and code notebook* (v1.1) [Computer software]. https://github.com/dalejn/cleanBib.

Zurn P, Bassett DS, Rust NC. 2020. The Citation Diversity Statement: A Practice of Transparency, A Way of Life. *Trends Cogn Sci* **24**:669–672.