

Adaptive Allocation of Vision under Competing Task Demands

Chris R. Sims, Robert A. Jacobs, and David C. Knill

Center for Visual Science, University of Rochester, Rochester, New York 14627

Human behavior in natural tasks consists of an intricately coordinated dance of cognitive, perceptual, and motor activities. Although much research has progressed in understanding the nature of cognitive, perceptual, or motor processing in isolation or in highly constrained settings, few studies have sought to examine how these systems are coordinated in the context of executing complex behavior. Previous research has suggested that, in the course of visually guided reaching movements, the eye and hand are yoked, or linked in a nonadaptive manner. In this work, we report an experiment that manipulated the demands that a task placed on the motor and visual systems, and then examined in detail the resulting changes in visuomotor coordination. We develop an ideal actor model that predicts the optimal coordination of vision and motor control in our task. On the basis of the predictions of our model, we demonstrate that human performance in our experiment reflects an adaptive response to the varying costs imposed by our experimental manipulations. Our results stand in contrast to previous theories that have assumed a fixed control mechanism for coordinating vision and motor control in reaching behavior.

Introduction

Nearly every human activity consists of a complex mixture of cognitive, perceptual, and motor processes. Although much is known about our perceptual and motor systems operating in isolation or in highly constrained tasks, less is known about the mechanisms that coordinate these systems in more complex settings. In this paper, we examine how gaze is controlled when the cognitive and motor systems place competing demands on vision. Such a scenario occurs when gaze must be distributed among two basic, yet important visual activities: using vision for closed-loop motor control, and for information acquisition to support planning future actions.

Previous research on the combined use of vision and motor control has focused on the role that vision serves in executing isolated motor tasks such as reaching to visually specified targets. A common finding is that, when visual feedback is available, the brain uses this information throughout the movement (Keele and Posner, 1968; Prablanc et al., 1979; Pélisson et al., 1986; Meyer et al., 1988; Abrams et al., 1990; Saunders and Knill, 2003; Heath, 2005).

Given the close relationship between vision and visually guided reaching, some have argued that the brain uses a common neural signal to coordinate these systems, or a “yoking” of eye and hand (Fisk and Goodale, 1985; Sailer et al., 2000). Evidence for this hypothesis comes from correlations among eye and hand

latencies in pointing tasks (Herman et al., 1981). Other evidence concerns the location of gaze during movements. In a series of experiments (Neggers and Bekkering, 2000, 2001, 2002), it was observed that the eye remained “anchored” at a target while performing a pointing task. The authors interpreted this as evidence that “saccadic execution is inhibited during goal-directed pointing movements” (Neggers and Bekkering, 2000). In support of the yoking hypothesis, Carey (2000) describes a patient who was unable to reach to targets not currently fixated, and often inappropriately reached to the point of fixation.

Common control of hand and eye movements represents a particularly simple solution to the visuomotor coordination problem: if the systems are yoked, then a mechanism for their independent control becomes unnecessary. A limitation of most previous research, however, is that the task is so severely constrained that execution of a movement becomes an end in itself, rather than a means to achieving behaviorally relevant goals. It therefore remains possible that evidence for a nonadaptive visuomotor coupling is an artifact of the tasks used to elicit behavior.

In the present research, we sought to study visuomotor coordination in a more natural task, in which we could independently manipulate demands on vision and examine resulting changes in behavior. Importantly, we also developed an “ideal actor” model that predicts the optimal coordination strategy. To preview our findings, we show that the timing of eye movements while reaching is not fixed but rather varies with changing task demands. The observed human performance is shown to be in close agreement with predictions from our ideal actor model.

Materials and Methods

Experiment 1

Overview. To investigate the coordination of visual gaze and motor control in interactive behavior, we designed a task that requires subjects to sort a series of rectangular shapes (“blocks”) according to their visual

Received Aug. 12, 2010; revised Oct. 6, 2010; accepted Oct. 26, 2010.

This work was supported by National Institutes of Health Grant R01-EY13319 (D.C.K.) and National Science Foundation Grant DRL-0817250 (R.A.J.). We thank Thomas Thomas for programming the experimental code and Leslie Chylinski for recruiting subjects and data collection.

Correspondence should be addressed to Chris R. Sims, Center for Visual Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627. E-mail: csims@cvs.rochester.edu.

DOI:10.1523/JNEUROSCI.4240-10.2011

Copyright © 2011 the authors 0270-6474/11/310928-16\$15.00/0

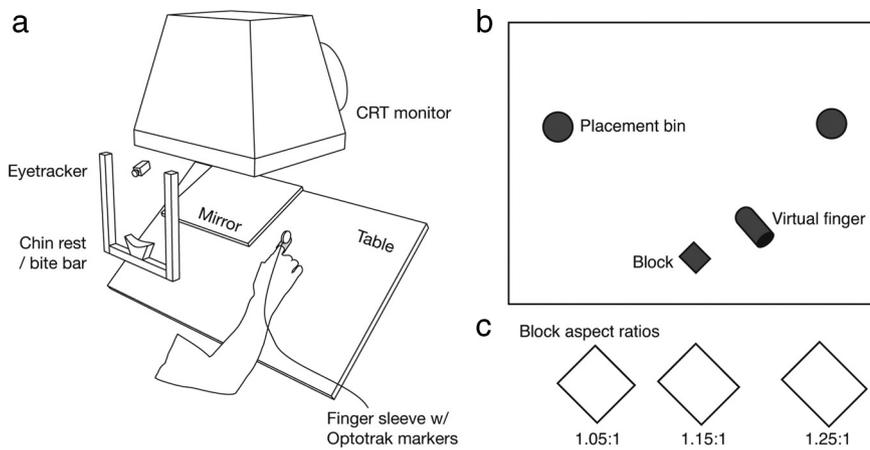


Figure 1. *a*, Experimental apparatus. Subjects interacted with a virtual workspace rendered in optical alignment with a physical table. The images were projected from an overhead CRT monitor and reflected toward the viewer using a half-silvered mirror. Subjects wore a metal sleeve with infrared markers on their right index finger. These markers were used to render a virtual fingertip in correspondence with the location of the subject's physical finger. *b*, Schematic diagram of the block-sorting task. Subjects had to sort blocks according to their orientation (counterclockwise in this example) and place them in one of two bins. In the actual task, stimuli were presented against a black background and there were no labels. *c*, Three levels of aspect ratio used in the experiment. In this example, all three blocks are rotated 45° counterclockwise.

appearance. The blocks consisted of rectangles that were rotated either 45° counterclockwise or 45° clockwise. Subjects had to pick up each block, and depending on the direction of its rotation, place it in one of two bins. Importantly, this simple task imposes competing demands on the visual system. On the one hand, visual gaze is needed to accurately determine the orientation of the blocks (using vision for information acquisition). At the same time, the motor act of picking up and accurately placing the blocks also requires visual guidance (using vision for on-line feedback control). We designed our task such that the difficulty of these two competing demands could be independently manipulated to investigate their effects on eye movements. How vision is divided between these two tasks determines how efficiently the task can be performed, and our goal was to examine whether subjects could optimally time their eye movements to maximize performance on the task.

We independently varied the demands that information acquisition and feedback control placed on the visual system. To manipulate the demands of on-line feedback control, we varied the size of the bins into which subjects had to place the blocks. Smaller bins require more precise motor control and should therefore require more visual guidance to accurately place the blocks in the bin. We therefore hypothesized that the demand placed on visual guidance should vary with the size of the target bins, with smaller bins requiring longer fixations on the placement bin, and later saccades away from the placement bin.

To manipulate the demands of information acquisition in the task, we varied the difficulty of the perceptual judgment that subjects had to perform to sort the blocks. In the experiment, subjects were required to sort rectangular blocks according to whether they were rotated clockwise or counterclockwise from vertical. In our experiment, the blocks were either rotated 45° counterclockwise or 45° clockwise. The difficulty of the perceptual judgment was manipulated by varying the aspect ratio of the blocks. By increasing the aspect ratio (making the rectangles more elongated), judgment of the direction of rotation becomes easier. Similarly, making the rectangles appear more square makes it more difficult to judge the direction of rotation. A harder perceptual judgment task should result in the observation of longer fixations on the blocks and earlier saccades to the blocks to allow the eyes more time for the perceptual judgment.

For a trial of a fixed duration, subjects must allocate some fraction of their visual gaze to guiding the hand to a target and some fraction of their gaze for information acquisition to plan the next movement in the task. If humans are able to strategically adapt their eye–hand coordination in response to varying demands of the task environment, then our manipulations on the bin size and block aspect ratio should produce observable

differences in the pattern and timing of eye movements. Importantly, we expected to observe anticipatory effects in the timing of eye movements: subjects should saccade toward the blocks sooner when the aspect ratio of the blocks makes perceptual judgment more difficult, and similarly, subjects should saccade away from the blocks sooner when the upcoming motor reaching task was more difficult. These results would indicate a strategic adaptation of visual allocation to the properties of the task environment. Furthermore, if the visual guidance system has evolved to support the efficient achievement of goals in interactive tasks such as this, then we might expect that, given sufficient experience in the task, subjects would adopt a nearly optimal allocation of their gaze.

Participants. Eight undergraduate students (six females) from the University of Rochester participated in the experiment. All subjects were right-handed and had normal or corrected-to-normal vision. All were naive to the purpose of the research. Subjects gave informed consent in accordance with guidelines from the University of Rochester Research Subjects Review Board.

Apparatus. Figure 1*a* shows a diagram of the experimental apparatus. Subjects viewed a virtual workspace that was projected from an overhead cathode ray tube (CRT) monitor and reflected through a half-silvered mirror. The monitor had a resolution of 1024 × 768 pixels and a refresh rate of 120 Hz. All stimuli were rendered in red to take advantage of the comparatively faster decay time of the red phosphor of the monitor. The mirror allowed the experimental software to render the virtual workspace in optical alignment with a physical table. The table was configured to a slant of 40° from horizontal such that the table surface was approximately perpendicular to the vector from the eye to the geometric center of the table.

Subjects wore a metal sleeve over their right index finger. An OptoTrak 3020 system recorded at 120 Hz the position of infrared markers mounted on this sleeve, and this position information was used to render in real time a virtual fingertip in correspondence with the subjects' true finger position. The OptoTrak system imposed a small latency on measurements of the finger position (~25 ms). To compensate for this delay, the rendered position of the virtual fingertip was linearly extrapolated ahead in time by 25 ms, using position data from recent frames to estimate finger velocity. The virtual finger was rendered as a cylinder with a rounded tip, with a radius of 1 cm and length of 5 cm. During the experiment, a matte black occluder was placed behind the mirror and prevented subjects from seeing their physical hand. The metal finger sleeve was also used to record when the finger made contact with the table: thin metal plates were mounted on the tabletop, and an analog-to-digital converter recorded when the metal finger sleeve made contact with the plates.

Subjects' gaze location was recorded during the experiment using an EyeLink II eyetracker (SR Research) operating at 250 Hz using corneal reflection. Subjects viewed the task monocularly using their left eye, and so only the left eye was tracked. To ensure accurate estimates of eye gaze, subjects' heads were held in place using a chin rest and bite bar.

To ensure an accurate perspective rendering of the virtual workspace and accurate data collection, each subject completed three calibration procedures before beginning the experiment. These procedures determined the physical position of the subject's eyes relative to the monitor (to ensure accurate perspective rendering), calibrated the position of the virtual fingertip to the subject's physical hand, and calibrated the eyetracker to ensure accurate gaze location data.

Stimuli and procedure. At the beginning of each trial, a crosshairs was displayed against a black background at the center of the workspace. Subjects began a trial by touching the center of the crosshairs. The pre-

cision requirements of touching this target were such that fixation on the crosshairs was generally necessary to successfully initiate a trial. As soon as the software detected that the subject's finger had made contact with the crosshairs, the display changed to show the main task. Figure 1*b* shows a schematic of this display.

The task display consisted of a rectangular block rotated to an orientation of either $+45^\circ$ or -45° from vertical, and two circular placement bins. The block was located 100 mm below the start crosshairs, where "above" and "below" to refer to the anterior and posterior direction in the frontoparallel plane of the work surface, respectively. The aspect ratio of the block varied according to experimental condition and was drawn from three levels: {1.05:1, 1.15:1, 1.25:1}. Figure 1*c* illustrates the three levels of aspect ratio used in the experiment. The length of the shorter side of the rectangle was fixed at 20 mm, whereas the longer side was determined based on the aspect ratio condition. The size of the placement bins also varied according to experimental condition, with radius equal to 8, 16, or 24 mm. The task for subjects was to decide whether the block was oriented to the left or to the right, and place the block in the appropriate bin: if the block was rotated counterclockwise, subjects had to place it in the left bin, and in the right bin if clockwise. Subjects touched a block to pick it up (the block "magnetically" attached to the finger) and then touched either the left or the right placement bin to drop it. To count as successfully placing a block, the measured location of the fingertip at the time of contact had to fall within the radius of the placement bin. The placement bins were located 200 mm from the pickup area (center-to-center distance). After touching the table to drop a block, the block disappeared from the end of the virtual fingertip.

Approximately 100 ms after picking up the first block, a second block appeared at the pickup location, regardless of whether subjects had placed the first block yet. The blocks always appeared at the same location in the workspace. To complete a single trial, subjects had to categorize and sort two blocks according to their orientation. Each block was randomly oriented to the left or to the right with equal probability. After placing the second block, the trial ended and subjects were shown a feedback screen. The feedback consisted of two icons that indicated the outcome for each of the two blocks placed during the trial. If a block was placed accurately in the correct bin, the icon was a green check mark. If a block was placed in the wrong bin, a double-headed arrow (\rightleftharpoons) was displayed. If the finger missed the placement bin, the icon was a red crosshairs. Finally, if the trial timed out before subjects could place a block, the feedback icon was an hourglass. Two icons were displayed on the feedback screen, one for each of the two blocks in a trial. Feedback was displayed for 1500 ms, at which point a crosshairs was displayed to begin the next trial.

Each subject completed 405 trials of the experiment in a single session lasting ~ 45 min. Each trial consisted of sorting and placing two blocks. The three levels of aspect ratio {1.05:1, 1.15:1, 1.25:1} and three bin sizes {8, 16, 24 mm} were crossed to produce nine within-subject conditions. For each experimental condition (defined by a combination of bin size and aspect ratio), subjects completed a run of trials. The order of the conditions was randomized. Completing a run of trials under all nine conditions defined an epoch of the experiment, and each subject completed five epochs.

The run length used was 5 trials per condition for the first epoch and 10 trials per condition for the remaining epochs. The first epoch was intended to give subjects practice on the task: subjects were given 10 s to complete each trial before they timed out. Thus, in a span of 10 s, subjects had to identify, pick up, and place both blocks. The trial time limit was reduced from 10 to 2 s for the last four "test" epochs of the experiment, so that subjects only had 2 s to place both blocks. In summary, each of the four test epochs consisted of 9 runs of trials; each run of trials used the same experimental condition, lasted for 10 trials, and the order of the runs was randomized.

After receiving verbal instructions on the apparatus and task, subjects were told that their goal was to maximize the percentage of trials that they completed correctly, where correct was defined as sorting and placing both blocks correctly within the trial time limit. Subjects were not given specific advice as to how to achieve this goal. Subjects were encouraged to take short breaks between each epoch.

Experiment 2

Overview. In a second experiment, we examined performance on just the perceptual judgment portion of the block-sorting task. A rotated block was presented on the screen, and subjects simply had to indicate whether the block was rotated counterclockwise or clockwise. The three levels of aspect ratio were the same as used in the first experiment. Rather than allowing subjects to freely view each block, we controlled the stimulus presentation duration using an adaptive staircase procedure (Levitt, 1971; Kollmeier et al., 1988). The purpose of this experiment was to quantify the demands that the perceptual judgment task placed on the visual system.

Participants. Nine undergraduate students (four females) from the University of Rochester participated in the experiment. All subjects had normal or corrected-to-normal vision. Five of the subjects had previously completed the first experiment, and the remaining four had not previously encountered the task or stimuli. Subjects gave informed consent in accordance with guidelines from the University of Rochester Research Subjects Review Board.

Apparatus. Subjects completed the experiment using the same display apparatus as in the first experiment (Fig. 1*a*); however, in the second experiment there was no motor component to the task. Consequently, subjects did not wear infrared markers on their hands and a virtual finger was not rendered in the workspace. Instead, subjects made responses by pressing either the left or the right button on a standard wireless mouse. Subjects viewed the display monocularly using their left eye.

Stimuli and procedure. On each trial of the experiment, a rectangle was briefly presented on the screen. The aspect ratio, size, and location of the stimuli in the display were identical with those used in the first experiment. After the stimulus presentation duration, a visual mask was immediately displayed on the screen for 500 ms. The mask consisted of a cluttered display of 500 overlapping and randomly oriented rectangles. Subjects then had 5 s to indicate their response. Subjects indicated a counterclockwise rotation by pressing the left mouse button, and a clockwise rotation by pressing the right mouse button. Visual feedback was then provided for 500 ms, indicating whether the subject had classified the previous trial correctly. The next trial was displayed immediately after the feedback duration. If subjects did not register a response within the 5 s response interval, the trial was repeated using a new (random) orientation.

The duration of the stimulus presentation was controlled using an adaptive staircase procedure (Levitt, 1971; Kollmeier et al., 1988). Three different staircase types were used for each of the three aspect ratios, and thus nine staircases in total were interleaved during the experiment. If a subject incorrectly classified a stimulus, the presentation duration for that staircase was increased by 8 ms. The staircases differed in terms of the number of correct responses in a row required to decrease the presentation duration for that staircase (also by 8 ms). The three staircase types used were one-up two-down, one-up three-down, and one-up four-down. The initial presentation durations chosen for these three staircases were 50, 250, and 500 ms, respectively. These parameters were chosen such that the resulting staircases collected data across a wide range of performance, from nearly chance level to nearly perfect performance. The nine staircases were randomly interleaved in the experiment, and subjects completed 225 trials per staircase, resulting in 2025 (225×9) total trials. The entire experiment was run in a single session lasting ~ 45 min.

Results

Experiment 1

The results reported in this section focus on subject performance during the final four test epochs of the experiment. All ANOVAs reported in this section are 3×3 repeated-measures ANOVAs with bin size ($r = \{8, 16, 24\}$) and block aspect ratio ($=\{1.05:1, 1.15:1, 1.25:1\}$) as factors.

Overall performance

On each trial, subjects were required to correctly place two blocks within a time limit of 2 s. On average, subjects placed

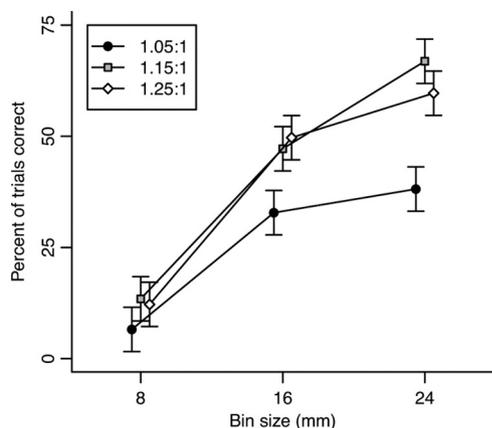


Figure 2. Percentage of trials completed correctly in each experimental condition. Error bars indicate 95% confidence intervals, computed using the method of Loftus and Masson (1994).

the first block correctly on 66% of trials; however, the percentage of trials on which subjects correctly placed both blocks dropped to 36%.

The percentage of trials completed correctly varied according to experimental condition, as shown in Figure 2. ANOVA on the percentage of trials completed correctly revealed that the main effects of both bin size [$F_{(2,14)} = 61.01$; mean square error (MSe), 1.26; $p < 0.001$] and aspect ratio ($F_{(2,14)} = 26.01$; MSe, 0.199; $p < 0.05$) were significant. As might be expected, performance increased as the size of the placement bins increased, and as the aspect ratio discrimination became easier. The interaction of bin size by aspect ratio was also found to be significant ($F_{(4,28)} = 6.04$; MSe, 0.029; $p < 0.05$). Subsequent t tests revealed that performance on the hardest and easiest aspect ratio conditions (1.05:1 vs 1.25:1) did not differ at the smallest bin size ($t_{(14)} = 1.44$; NS), but this performance difference was significant by the largest bin size ($t_{(14)} = 2.43$; $p < 0.05$), resulting in the significant interaction. Inspection of Figure 2 also reveals that the difference in performance between the two easiest aspect ratios was negligible compared with the difference between the moderate and hardest aspect ratios.

Motor performance

In this section, we focus on two measures of motor behavior: the movement durations when reaching to pick up or place the blocks, and the contact duration of the finger while picking up and placing blocks.

The first measure of motor behavior we examine is the movement duration for the four motor segments of the block-sorting task: moving the hand to pick up the first block (pickup-1), the movement from the pickup area to the placement bin (place-1), returning the hand to pick up the second block (pickup-2), and movement of the hand to place the second block (place-2). Movement time was defined as the time interval from when the finger left contact with the table (after, e.g., picking up a block), to the time when the finger again made contact with the table (when placing the block). Movement times were only recorded for complete motor segments, so, for example, if the trial timed out while the subject was reaching to place the second block, no movement time was recorded for this movement segment. Movement time outliers differing by > 2 SDs from the mean were removed before performing analyses.

Figure 3 reports the mean movement times for each experimental condition. For the initial movement from the start crosshairs to pick up the first block, ANOVA revealed a significant

effect of aspect ratio ($F_{(2,14)} = 8.57$; MSe, 5962.2; $p < 0.05$). Subsequent paired t tests revealed that subjects' mean movement time in the 1.05:1 aspect ratio condition was significantly slower than in the 1.15:1 condition ($t_{(7)} = 5.23$; $p < 0.05$), whereas there was no significant difference between the 1.15:1 and 1.25:1 conditions. Thus, subjects were significantly slower in picking up the first block when the aspect ratio judgment was hardest, compared with the two easier aspect ratio conditions.

Although aspect ratio significantly influenced the movement time to pick up the first block, the movement time associated with placing that block showed a different pattern of results. ANOVA revealed no significant effect of aspect ratio, but instead a main effect of bin size ($F_{(2,14)} = 18.04$; MSe, 18,933.2; $p < 0.05$): subjects were slower in moving their hand to place a block in a smaller bin.

Interestingly, the bin size of the targets also influenced how long it took subjects to pick up the second block after placing the first ($F_{(2,14)} = 11.97$; MSe, 9810.0; $p < 0.05$). While placing the blocks, a slower movement time toward smaller targets is expected based on Fitts's law (Fitts, 1954). However, the effect of bin size on movement time to pick up the second block is surprising. In essence, subjects were slower to pick up the second block after having just placed the first block in a small bin, compared with the case of placing the first block in a larger bin.

The movement duration while placing the second block revealed a similar pattern of results to placing the first block. Movement duration was not influenced by aspect ratio, but there was a main effect of bin size ($F_{(2,14)} = 11.0$; MSe, 4728.4; $p < 0.05$). As before, subjects were slower in placing the block in a smaller bin, and the effect of aspect ratio was not significant.

In summary, subjects took longer to pick up the first block as the aspect ratio decreased. After picking up this block, the aspect ratio had no influence on movement durations for the rest of the trial. The bin size influenced movement duration for placing both blocks, and also influenced how long it took to pick up the second block.

Contact duration was measured using the metal sleeve worn over the subject's finger. Across all conditions, the mean contact duration while picking up a block was 145 ms, and the mean contact duration while placing the first block was 88 ms. Contact duration while placing the second block is not well defined in our experiment, as the trial ends as soon as contact is detected on placing the second block. While picking up the first block, contact duration varied significantly depending on the aspect ratio ($F_{(2,14)} = 11.52$; MSe, 546.66; $p < 0.05$). Subsequent paired t tests revealed that contact duration was longer for the 1.05:1 aspect ratio condition than for 1.15:1 ($t_{(7)} = 3.56$; $p < 0.05$), but there was no difference between the two harder aspect ratios. Conversely, when placing the first block, there was an effect of bin size on contact duration ($F_{(2,14)} = 13.00$; MSe, 3633.4; $p < 0.05$), but no effect of aspect ratio. Subjects demonstrated longer finger contact durations when placing blocks in smaller bins compared with larger bins. Curiously, while picking up the second block, there was also a significant effect of bin size on contact duration ($F_{(2,14)} = 9.91$; MSe, 335.96; $p < 0.05$): contact duration was significantly longer when subjects had previously placed a block in a small bin, compared with either of the larger bin sizes. This carryover effect parallels the results previously presented for movement durations and will be discussed further in conjunction with our ideal performer model.

Eye movements

Subjects' gaze position was monitored throughout each trial. The raw data from the eyetracker consisted of the estimated gaze po-

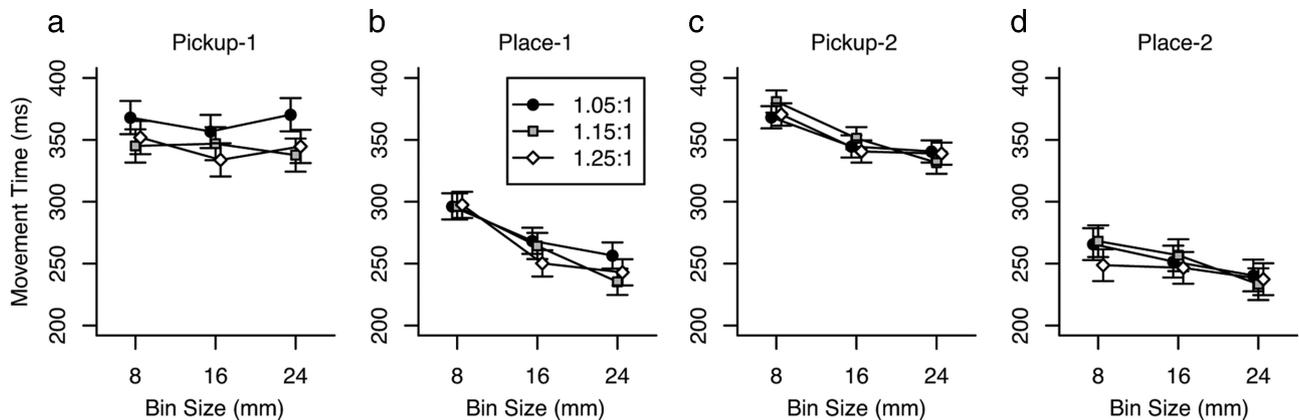


Figure 3. Movement durations for all four motor segments that comprised each trial. *a*, Duration of reach to pick up the first block. *b*, Duration of movement to place the first block. *c*, Duration of movement to pick up the second block. *d*, Duration of movement to place the second block. Error bars indicate 95% confidence intervals, computed using the method of Loftus and Masson (1994).

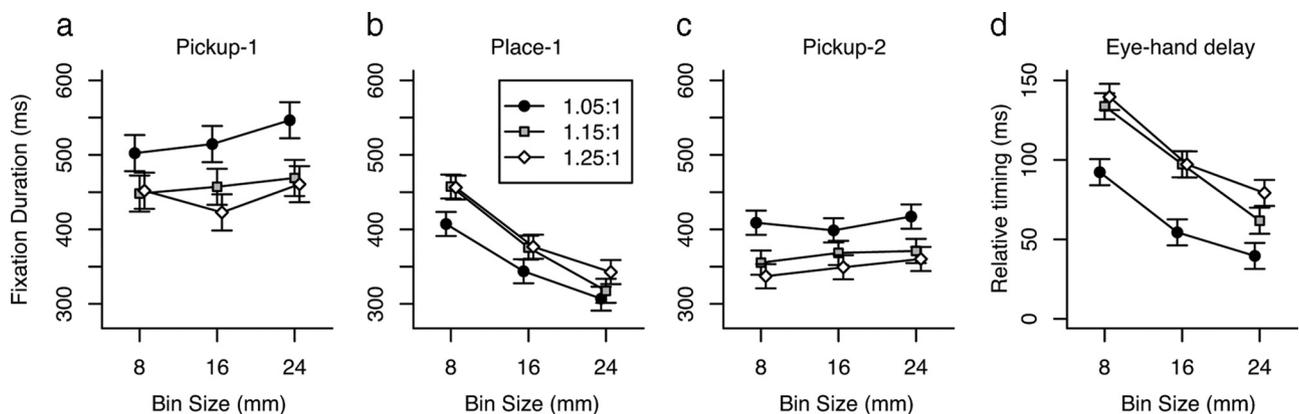


Figure 4. Eye movement data. *a*, Mean duration of fixation while picking up the first block. *b*, Mean duration of fixation on the placement bin while placing the first block. *c*, Mean fixation duration while picking up the second block. *d*, Eye–hand delay, defined as the time of the first saccade from the placement bin back to the pickup area, relative to when the finger arrives at the placement bin to place the first block. Note the differing axis for *d*. Error bars indicate 95% confidence intervals.

sition on each time frame, sampled at 250 Hz. These data were segmented into a series of fixations and saccades in the following manner. First, a fourth-order polynomial was fit to a sliding window of five data points. The third temporal derivative of this polynomial (jerk) was used to detect the presence of a saccade, using a threshold of 2 mm/frame³. The time of saccade onset was determined as the first sample above this threshold. Saccade termination was determined as the first sample after saccade onset that fell below a velocity threshold of 5 mm/frame. The extracted saccades were used to segment the eye data into a series of fixations, with the fixation location computed as the mean gaze location for all data assigned to that fixation. Each fixation was subsequently assigned to a destination on the work surface (either the start cross, pickup area, or placement bin) if the mean gaze location of the fixation fell within 5 cm of the center of the relevant target. If the fixation could not be assigned to one of these locations, it was classified as “NA” and excluded from the analysis.

In this section, we will focus on four measures of gaze behavior. Pickup-1 fixation duration (Fig. 4*a*) is defined as the mean duration of fixation on the pickup area while picking up the first block. This includes all eye fixations on the block that occurred after the start of the trial and before making any saccades to the placement bin. ANOVA revealed significant main effects of both bin size ($F_{(2,14)} = 4.45$; MSe, 5963.4; $p < 0.05$) and aspect ratio ($F_{(2,14)} = 12.5$; MSe, 41,015.0; $p < 0.05$). Subjects fixated the block longer when the aspect ratio was harder, compared with the

case when the aspect ratio was one of the two easier conditions. At the same time, subjects spent less time fixating the block when they subsequently had to place it in a smaller bin. This latter finding suggests that subjects adaptively anticipated the difficulty of an upcoming motor plan: by shortening the duration of fixation on the blocks, subjects allocated more time for vision to guide the hand to the smaller placement bins.

The second measure of gaze behavior, place-1 fixation duration (Fig. 4*b*) is defined as the duration of fixation on the placement bin while placing the first block (all fixations from the first saccade toward the placement bin, until the first saccade back toward the pickup area). ANOVA indicated significant main effects of both bin size ($F_{(2,14)} = 24.87$; MSe, 82,389.0; $p < 0.05$) and aspect ratio ($F_{(2,14)} = 4.16$; MSe, 7456.1; $p < 0.05$). Subjects spent more time fixating the placement bin as the size of the placement bin decreased. At the same time, subjects adaptively spent less time fixating the placement bin as the difficulty of the perceptual judgment increased. These data are consistent with the idea that subjects knew that performing the perceptual judgment in the harder aspect ratio condition would require longer fixations and therefore planned the timing of their eye movements accordingly.

The third measure of gaze behavior is the duration of fixation on the pickup area while picking up the second block (pickup-2 fixation duration) (Fig. 4*c*). ANOVA revealed a significant main effect of aspect ratio ($F_{(2,14)} = 11.30$; MSe, 19,623.8; $p < 0.05$); as

with picking up the first block, subjects spent longer fixating the block when the aspect ratio was drawn from the hardest condition, compared with the two easier aspect ratio conditions. Unlike the first block, however, the main effect of bin size while picking up the second block did not reach significance ($p > 0.05$).

The previous three measures of gaze behavior have looked at the duration of fixations on the blocks and placement bin. In addition to gaze duration, another important feature of human performance in this task is the relative timing of eye movements compared with the timing of hand movements. A particularly important question regarding the relative timing of eye movements is when the eye saccades back to the pickup area to fixate the second block, relative to when the hand first makes contact with the work surface to place the first block. Previous research investigating look-ahead fixations (Pelz and Canosa, 2001; Pelz et al., 2001; Mennie et al., 2007) suggests the possibility that, in some conditions of the experiment, subjects would initiate a saccade to determine the orientation of the second block, even before the hand has completed placing the first block.

Figure 4*d* plots the timing of the saccade to the second block relative to the time that the hand places the first block, a measure which we term “eye–hand delay.” An ANOVA on eye–hand delay revealed significant main effects of both bin size ($F_{(2,14)} = 31.07$; MSe, 24,066.0; $p < 0.05$) and aspect ratio ($F_{(2,14)} = 9.50$; MSe, 9963.2; $p < 0.05$). The eye departed the placement bin sooner as the bin size increased; the eye also departed sooner as the difficulty of the upcoming perceptual judgment increased. Note that, for all conditions, the eye–hand delay was positive, which indicates that saccade onset occurred after the finger made contact with the work space to place the first block. In other words, look-ahead fixations were not observed in our task. However, the shortest eye–hand latency observed was on the order of 40 ms. Research on the time course of saccade generation indicates that motor planning processes in the ocular system require on the order of 250 ms (Becker, 1991), and thus there is strong evidence that the observed saccades toward the pickup area were planned well before the finger made contact with the placement bin. This result is of importance, as previous research (Neggers and Bekkering, 2000) has claimed that both saccade execution and saccade planning are inhibited during the course of manual pointing movements. Finally, the ANOVA also indicated that the interaction between bin size and aspect ratio was significant ($F_{(2,14)} = 5.2497$; MSe, 674.87; $p < 0.05$). Inspection of Figure 4*d* reveals the origin of this significant interaction; eye–hand delay did not differ among the two easier aspect ratio conditions at the smaller bin sizes ($r = 8, 16$), but this difference was significant for the largest bin size.

Summary

Many of the empirical results conform to predictions that would hold if subjects planned their visuomotor behavior for the two components of the task independently. Subjects took longer to move their hand to a target when placing objects in smaller targets, as would be expected based on Fitts’s law (Fitts, 1954), and also fixated those targets longer. Similarly, subjects fixated the blocks at the pickup area longer when the aspect ratio made the perceptual judgment more difficult. The interesting results from this experiment are those that demonstrate adaptive coordination of visual gaze and motor control among the two components of the task. One critical result is that subjects spent less time determining the orientation of a block if they had to place it in a smaller bin. In other words, the difficulty of the motor task im-

acted the amount of vision allocated to perceptual judgment. This effect was observed even on the first block of a trial, ruling out the possibility that a more difficult motor task interfered with a subsequent perceptual judgment. Instead, it appears that subjects adaptively planned the duration of their initial fixation based on the difficulty of the upcoming motor task.

A second critical finding from this experiment is that subjects spent less time fixating the placement bins if they had to perform a harder aspect ratio discrimination on the next block. As before, an interference-based account can be ruled out: It is not the case that the eyes arrived later on the placement bins after a harder orientation judgment, resulting in shorter fixations on the placement bin. Instead, as Figure 4*d* illustrates, subjects made a saccade away from the placement bin sooner during trials with a difficult aspect ratio condition.

Depending on the experimental condition, the timing of eye movements relative to the hand varied by as much as 100 ms, or approximately one-third of the typical duration of a reaching movement in the experiment. These results demonstrate that the timing of saccades in this experiment reflected an adaptation to the demands of the task. Furthermore, these results cannot be explained by theories of eye–hand coordination that assume that the eye and hand are yoked with regard to the onset of movement or timing of saccades.

There is one result from the experiment that cannot intuitively be explained by an adaptive coordination of the visual and motor systems. Namely, the size of the bin while placing a block significantly influenced the duration of the motor movement after placing that block (Fig. 3*c*). In effect, subjects exhibited a correlation in the movement durations for the two motor segments when there was no apparent reason to do so. We will consider possible explanations for this finding in conjunction with our ideal actor model and in Discussion.

One additional finding emerges from examination of the eye movement data. For all four measures of eye gaze considered (Fig. 4*a–d*), there was a relatively large difference in gaze behavior between the hardest aspect ratio condition and the medium difficulty condition, but little or no difference in gaze behavior between the medium and easiest condition. Interestingly, this pattern of results mirrors the data on task accuracy (Fig. 2), in which there was a significant difference in percentage of trials correct for the hard versus medium aspect ratio, but little difference in performance between the medium and easy aspect ratio condition. Although not conclusive, this parallel suggests that the nature of the observed changes in saccade timing were closely tied to their impact on performance in the task. To investigate this possibility further, our second experiment was conducted to specifically examine the relationship between viewing time of the oriented blocks and accuracy in judging the direction of rotation.

Experiment 2

In our second experiment, we examined performance on just the perceptual discrimination portion of the block-sorting task. Performance was analyzed by fitting to each subject’s entire history of choice data a model that predicts probability of correct judgment as a function of stimulus presentation duration and aspect ratio. In recent years, numerous researchers have studied the trade-off between response speed and accuracy in perceptual judgments (for review, see Bogacz et al., 2010). A common assumption among models is that, at some processing level, decision making involves accumulating noisy or uncertain evidence over time. One such model incorporating this assumption is known as the Wiener diffusion model (WDM) (Ratcliff and

Smith, 2004; Zhang et al., 2009). The WDM has been able to account for a wide range of empirical findings in two-alternative forced-choice tasks (Ratcliff and Rouder, 1998), including perceptual judgments. As applied to our experiment, the model assumes that subjects accumulate evidence regarding the orientation of a block over the course of the stimulus presentation interval. The parameter of interest is τ , or how quickly subjects accumulate evidence to support their decision. In our task, we are interested in estimating three separate parameters τ , one for each of the three aspect ratio conditions, under the hypothesis that subjects will accumulate evidence more slowly as the aspect ratio decreases. Under the assumptions of the WDM, the probability of correctly judging the orientation of a block varies as a function of viewing duration t and evidence accumulation rate τ according to the following:

$$p(\text{correct}) = \frac{1}{2} \left(1 + \text{Erf} \left(\frac{\tau}{\sigma} \cdot \sqrt{\frac{t}{2}} \right) \right), \quad (1)$$

where Erf indicates the Gaussian error function. Note that Equation 1 also includes a stimulus noise parameter, σ . However, since only the proportionality of τ and σ matters for the predictions of the model, we arbitrarily set $\sigma = 1$ and estimated the remaining rate parameter. Equivalently, we could have fixed τ and estimated a separate noise parameter for each aspect ratio condition. The model was fit to each subject's data separately using maximum-likelihood estimation (Myung, 2003). For each subject, the data consisted of the presented stimulus duration, and whether the subject judged the orientation correctly. Let δ_i be an indicator variable that takes the value 1 if the subject responded correctly on trial i , and 0 otherwise. Similarly, let t_i indicate the stimulus duration on trial i . Then the likelihood function for all trials in a given aspect ratio condition is given by the product of the probabilities of the subject being correct on each trial:

$$\prod_{i=1}^N [\delta_i \cdot p(\text{correct}|\tau, t_i) + (1 - \delta_i) \cdot (1 - p(\text{correct}|\tau, t_i))]. \quad (2)$$

This likelihood function was maximized by fitting the evidence accumulation rate for each aspect ratio condition. The numerical maximization procedure was repeated several times using different initial parameters, and the best-fitting parameters were retained for each subject. We also examined two alternatives to the WDM, including a version with a lapse rate parameter, and the Ornstein–Uhlenbeck model (Ratcliff and Smith, 2004; Zhang et al., 2009), which incorporates an additional evidence decay parameter. It was found that neither of these variations improved the fit of the model to the data.

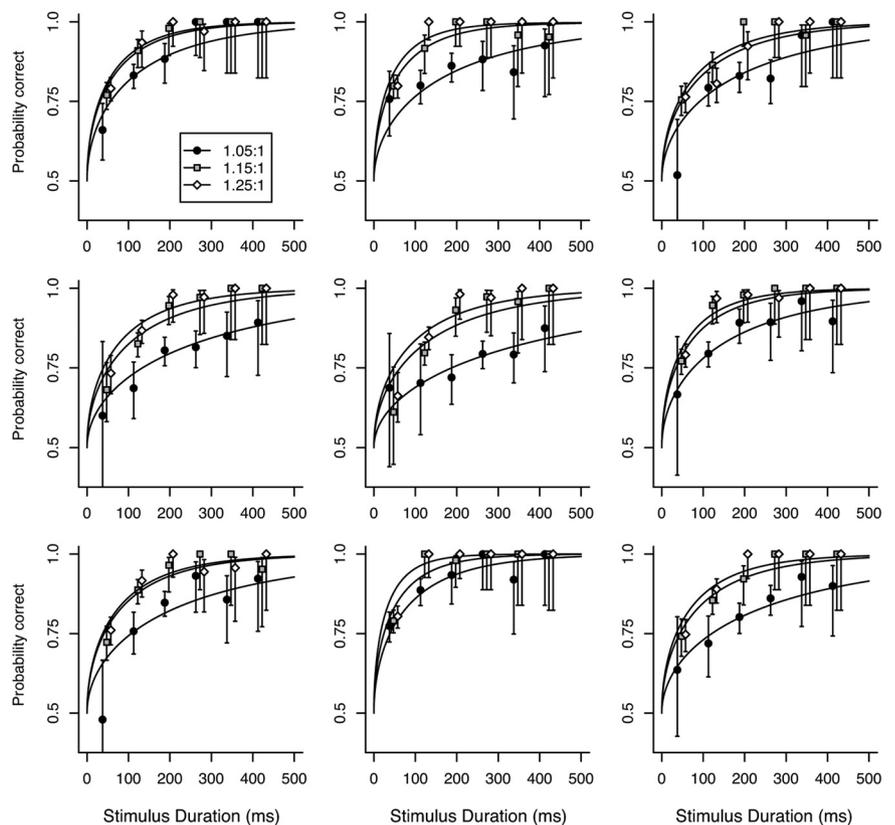


Figure 5. Probability of correctly judging block orientation as a function of stimulus presentation duration and aspect ratio. Each panel shows the best fitting model for an individual subject. The smooth curves indicate model predictions. The markers indicate mean subject performance, after trials were grouped into bin sizes of 75 ms. Error bars indicate the Bayesian 95% credible intervals, assuming a uniform prior over probability correct for each bin.

Figure 5 shows the resulting maximum-likelihood model fits for each of the nine subjects, plotting the probability of correctly judging the orientation of a block as a function of stimulus presentation duration and aspect ratio. The human data are indicated by the markers, in which trials have been grouped into bins of size 75 ms. Binning the human data was necessary as the staircase procedure varies the stimulus duration from trial to trial. Note, however, that the models were fit to the trial-by-trial responses, and not the binned data.

As shown in Figure 5, all nine subjects demonstrated a similar trade-off between viewing time and accuracy. Accuracy improved more quickly for the easier aspect ratio conditions. Furthermore, there are no substantive differences between the performance of subjects who had previously completed the first experiment (panels 1–5) and novice subjects (panels 6–9). Using the model predictions for each subject, we computed the 90% threshold values, or the stimulus duration at which subjects reached 90% correct classification for each aspect ratio. Paired t tests revealed that the difference in 90% threshold between the aspect ratio conditions was significant. Comparing the 1.05:1 and 1.15:1 aspect ratio conditions, the mean difference was 216.10 ms ($t_{(8)} = 5.471$; $p < 0.05$). For 1.15:1 versus 1.25:1, the mean difference was 23.36 ms ($t_{(8)} = 2.95$; $p < 0.05$).

In summary, in this experiment, subjects performed the same perceptual judgment task as the first experiment—namely, judging the direction of oriented rectangles—with the exception that stimulus duration was experimentally controlled rather than implicitly governed by subjects' own eye movements, and there was no simultaneous motor task to be

performed. In experiment 2, subjects required longer stimulus presentation durations to reach the same level of accuracy when the difficulty of the perceptual judgment increased.

Although this basic result is not surprising, the relative magnitude of the difference between the conditions is interesting in light of the results obtained in the previous experiment. Previously, it was observed in experiment 1 that subjects initiated a saccade from the placement bin to the second block in a trial sooner when that block had a harder aspect ratio (Fig. 4*d*). In that experiment, the difference between the hardest and medium aspect ratio conditions was relatively large, whereas there was no significant difference found between the two easier conditions. In experiment 2, a similar pattern emerged in terms of the relative amount of perceptual input required to achieve a criterion level of performance in the three aspect ratio conditions. A large difference was found between the hardest and medium difficulty conditions, and only a small difference between the two easier conditions. These results indicate that not only did subjects adaptively plan the timing of their eye movements in the block-sorting task but did so in a manner that was remarkably sensitive to the low-level properties of the task; namely, the relationship between the duration that the blocks were viewed and the accuracy at judging their orientation.

In the next section, the quantitative results of the first two experiments will be used to form an ideal actor model for the combined block-sorting task.

Ideal actor analysis

In this section, we derive a model that predicts the optimal coordination of eye and hand movements in the block-sorting experiment. Our model falls in the family of ideal actor analyses developed in other research contexts (Chhabra and Jacobs, 2006; Gray et al., 2006). By comparing human performance to a model of optimal performance, several advantages are gained. First, if humans are found to behave in an optimal or near-optimal manner, then the observed behavior can be understood as a rational adaptation to the demands of the task and constraints on the human system, rather than the by-product of ad hoc or arbitrary mechanisms whose existence is postulated solely to account for the observed data. Second, the finding of adaptive behavior in a task rules out the possibility that the brain uses fixed, task-independent control mechanisms to govern behavior. Third, an ideal actor model enables an understanding of the observed behavior (in terms of rationally achieving task-relevant goals) even if the biological mechanisms responsible for producing the observed behavior are not fully understood. Finally, the finding of optimal behavior suggests that humans are using all available information in an efficient manner; this places strong constraints on the neural mechanisms necessary to achieve such performance.

In developing our ideal actor model, we found it necessary to model just a portion of the complete block-sorting task, to limit the number of assumptions in our model, as well as to reduce the computational complexity of performing simulations. Our model simulation begins at the moment the subject has just picked up the first block in a trial, as the hand leaves the pickup area to place this block. Our model concerns the motor act of placing this block, returning the hand to pick up the second block, and the concurrent perceptual judgment of the orientation of this second block (thus, we model one “round” of the block-sorting task, in which a single trial consists of two rounds).

Our ideal actor model for the block-sorting task consists of model visual and motor systems that incorporate the important

constraints of these systems in defining optimal task performance. In particular, the model of the motor system incorporates biologically realistic motor noise, such that executing rapid reaching movements is inherently noisy and error prone. Visual feedback is needed to accurately control the motor system in an on-line fashion. Optimal motor control signals are generated using stochastic optimal feedback control theory (Todorov and Jordan, 2002; Todorov, 2004; Diedrichsen et al., 2010), and our model of the visual system incorporates time-delayed and noisy sensory integration. Vision for the perceptual judgment of block orientation is also modeled in a realistic manner, using the data from our second experiment to directly constrain the performance of the model.

Since our complete model is rather complex, we briefly describe the motor component and the visual component of the model and provide a more detailed description of each in Appendix. Then, the costs and constraints on performance for the model are specified, before discussing the predictions of the model.

Modeling the motor system

We chose to adopt a simplified model of the motor system that nonetheless retains the important characteristics of human performance in the block-sorting task. Following Liu and Todorov (2007), we model the hand as a point mass moving in a two-dimensional plane. The hand is controlled by forces that act in orthogonal directions; these forces are subject to low-pass filters that approximate the properties of human muscle. Furthermore, the generated forces are corrupted by biologically realistic multiplicative noise, such that larger control signals result in greater variability of the resulting motor execution. Such control-dependent noise has previously been shown to be critical in accounting for the smooth velocity profiles observed in human reaching movements (Harris and Wolpert, 1998).

In implementing the model, it is useful to define a state vector $\mathbf{x}(t)$ that contains all the variables that describe the current state of the hand. In our case, this state vector includes the position, velocity, and muscle state of the hand, as well as the locations of various objects in the workspace (see Appendix). The dynamics of this state vector can then be described by the following general discrete system:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \xi_t + \sum_{i=1}^c \varepsilon_t^i C_i \mathbf{u}_t. \quad (3)$$

The matrices A and B are derived from the continuous-time dynamics of the hand, described more fully in Appendix. The vector ξ_t is additive Gaussian noise, and the last term specifies the multiplicative Gaussian noise model.

Generally speaking, the task for the motor system is to produce the time-varying control signals \mathbf{u}_t such that the hand accurately and quickly moves from the pickup location to the placement bin to place the first block and returns to the pickup location to pick up the second block. The optimal control signals for accomplishing this will depend on the current estimates of the system state; these estimates will in turn be based on noisy and time-delayed information conveyed by the sensory system.

Vision for on-line feedback control

In the block-sorting task, as in many natural tasks, vision serves multiple roles in supporting efficient performance. First, vision is useful for guiding hand movements via on-line feedback control. Second, in our task visual gaze is necessary for determining the

orientation of the blocks that are to be sorted on each trial. We describe our model of visual processing for each of these two roles in turn. Additional details of the sensory model are provided in Appendix.

In the normal course of reaching movements, the visual system integrates information about the position and velocity of the hand, as well as information about the position of the target that the hand is reaching for. To incorporate these properties of the human visual system, we implement our model as follows. Given the state of the system \mathbf{x}_t at time t , the model receives sensory information regarding a subset of these variables, \mathbf{y}_t . This sensory information is degraded by both additive and multiplicative noise. This yields a sensory model of the following form:

$$\mathbf{y}_t = H\mathbf{x}_t + \omega_t + \sum_{i=1}^d \varphi_i^j D_i \mathbf{x}_t, \quad (4)$$

where ω represents additive Gaussian noise and the term involving D_i represents the multiplicative component of the sensory noise (φ is zero-mean, unit variance Gaussian noise). Briefly, our sensory noise model incorporates effects of retinal eccentricity on estimating position, as well as effects of velocity on sensory noise. In addition, the sensory observation model defined by Equation 4 was extended in our implementation to incorporate time-delayed feedback, as well the effects of saccadic suppression (Bremmer et al., 2009). Each of these components of our model is based on known psychophysical limits, and a more detailed account of the sensory model is provided in Appendix.

Given time-delayed and noisy sensory information, our ideal actor model optimally integrates this information to produce a “best estimate” of the current state, $\hat{\mathbf{x}}_t$. For systems with linear dynamics and Gaussian noise, the form of this estimate is given by the well known Kalman filter (Kalman, 1960) as follows:

$$\hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t + B\mathbf{u}_t + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_t) + \eta_t. \quad (5)$$

The matrix K_t specifies the time-varying Kalman gains for combining incoming sensory signals with the current estimate of the state. These gains are chosen to minimize performance costs in executing reaching movements (these costs are discussed in the next section). The final term in the estimator, η_t , represents Gaussian noise, or random drift in the state estimate of the model. When the visual system is attending to the placement bin, the model does not integrate sensory information about the block pickup location (and vice versa), and therefore the estimate of this location decays or degrades over time. This drift in the internal estimate was necessary in our model as otherwise the model would be able fixate the pickup location at the beginning of the experiment and maintain a perfect representation of this location without ever having to fixate it again, clearly in contradiction to realistic limitations of human visual short-term memory.

Vision for information acquisition

In addition to guiding the hand during reaching movements, the visual system must also be used for gathering information about the orientation of each block that has to be sorted. The important characteristic of the visual system for our model is the relationship between fixation duration and accuracy in judging the orientation of the blocks. Since our second experiment provided direct evidence for this relationship, we simply use the Wiener diffusion model fit to the human data from our second experiment to constrain the visual discrimination performance of our model. As all nine subjects in the second experiment demon-

strated similar perceptual discrimination performance (Fig. 5), we averaged the best-fitting parameters for each subject to produce a model of the mean discrimination performance as a function of aspect ratio and viewing duration.

In our model, evidence regarding the orientation of the next block begins accumulating 100 ms after the model makes a saccade to the pickup location because of the sensory delay. Evidence continues to accumulate from this point until the finger makes contact to pick up the block. At this point, our model assumes that the subject has committed to placing the block in one of the two bins. The amount of evidence that the model has available for the orientation discrimination task therefore depends on the timing of its eye movements; this in turn trades off with the amount of time that the eye can spend fixating the placement bin to accurately guide the hand when placing the previous block.

Cost function on behavior

Any notion of optimality must be defined relative to some cost that is to be minimized, or equivalently a utility function that is to be maximized. In our model, we have assumed a two-level hierarchical cost function on performance. At the top level of this hierarchy, high-level kinematic parameters of behavior are programmed to maximize task performance. These parameters include the durations of motor movements—reaching from the pickup area to the placement bin and back again—as well as the timing of eye movements between the pickup and placement locations. These kinematic parameters are chosen to optimize performance on the task, in which at this level performance is defined as maximizing the percentage of trials that are completed correctly within the trial time limit. In our model, this is defined as the combined success rate of three task components: accurately placing the first block in the bin, accurately touching the second block to pick it up, and accurately judging its orientation.

At the lower level of the control hierarchy, the motor system treats the specified movement durations and eye movements as constraints and optimizes on-line motor performance subject to these constraints. For a given movement duration and visual gaze allocation, the motor system must determine the optimal sequence of control signals \mathbf{u}_t and Kalman filter gains K_t that result in motor endpoints that are as accurate as possible. Previous research on the biological cost function for motor control suggests that an appropriate cost can be defined in terms of a quadratic penalty on endpoint error (Körding and Wolpert, 2004), although for large errors this approximation may be inaccurate. Since our state vector includes both the hand position as well as the target locations, we can define the motor cost function as a quadratic function of the state vector at each time step t as follows:

$$\text{Cost per step} = \mathbf{x}_t^T Q_t^x \mathbf{x}_t + \mathbf{u}_t^T Q^u \mathbf{u}_t. \quad (6)$$

In this equation, Q_t^x is a matrix that specifies the quadratic cost on the state variables at time t , whereas Q^u specifies the cost on the magnitude of the control signals applied to the hand. In our implementation, Q_t^x specifies a quadratic penalty on the difference between the hand and target position at the end of each movement segment. Since the hand must also maintain stability while picking up and placing the blocks, this matrix also includes a cost on the velocity of the hand at the end of the movement. Empirically, subjects' fingers remained in contact with the work surface for ~ 100 ms when picking up and placing the blocks. We therefore imposed a quadratic cost on velocity for the empirical contact durations at the end of each movement segment. The

costs on the state vector were set to zero during the course of the movement, so that only endpoint costs were specified.

The exact magnitudes of the various cost terms does not matter, but only their relative magnitude. We therefore set the cost on the positional error of the hand at the end of a reaching movement to 1; the costs on terminal velocity of the hand and the control signal costs were chosen to produce realistic motor trajectories and were set to 1.0×10^{-3} for velocity and 7.0×10^{-6} for control cost.

For a linear system as specified in Equation 3, with an observation model as in Equation 4, a linear estimator as shown in Equation 5, and a cost function given by Equation 6, the task for the motor control system is to derive the time-varying motor commands \mathbf{u}_t and filter gains K_t that minimize the total expected cost. When only additive Gaussian noise is present, the solution to this problem can be calculated analytically. In the presence of multiplicative noise, no closed-form solution is known to exist. However, Todorov (2005) has presented an iterative numerical algorithm for efficiently computing the optimal controller and estimator for this system. In this work it is shown that the optimal motor control signals, \mathbf{u}_t^* , take the form of a linear feedback control law based on the current estimate of the system state as follows:

$$\mathbf{u}_t^* = -L_t \hat{\mathbf{x}}_t, \quad (7)$$

where the time-varying feedback gains L_t can be computed in advance of the movement.

To summarize the development thus far, our ideal actor model consists of a model of the motor and visual systems. The visual system is responsible for both providing on-line feedback control to the hand and for determining the orientation of the blocks that have to be sorted. Optimality is defined with respect to a two-level hierarchy. First, high-level kinematic parameters of behavior are specified to maximize task performance. These parameters govern the durations of motor movements and the timing of saccades. For a given set of high-level parameters, the motor system treats these parameters as constraints and optimizes motor accuracy subject to these constraints. At this lower level of the hierarchy, performance is defined by minimizing quadratic penalties on inaccuracy.

The goal of the ideal actor analysis is to determine the visuo-motor coordination strategy that results in optimal performance on the task. The optimal performance is then compared with human behavior for evidence of optimality in human visuo-motor coordination in this task.

Model results and predictions

In our first experiment, we observed that the eye always made a saccade to the placement bin before the finger left contact with the pickup area. Therefore, at the start of the simulation the hand position was initialized to the block pickup location, and gaze was initialized to the placement bin. The high-level control parameters of our model include the movement duration in reaching to place the first block (place-1 movement duration), reaching to pick up the second block (pickup-2 movement duration), and the duration of the fixation on the placement bin while placing the first block (place-1 fixation duration). At the end of this fixation, the model executes a saccade to the pickup location to guide the hand in picking up the next block and to judge the orientation of this block. As a first test of the predictions of the model, we constrained the model to use the empirically observed movement and finger contact durations. With these motor parameters fixed, a single parameter, place-1 fixation duration, remained. We op-

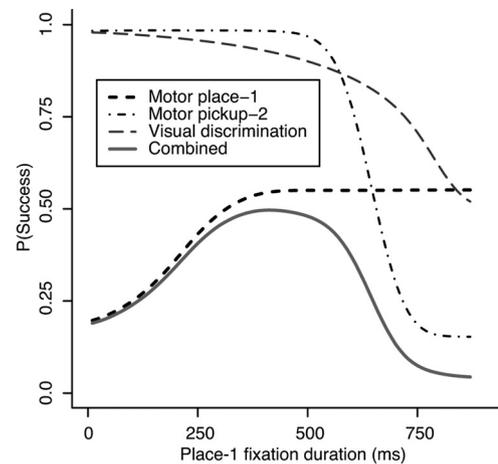


Figure 6. Example illustrating the calculation of the utility function predicting the optimal timing of eye movements. The predicted utility is the joint probability (i.e., the product) of correctly placing the first block, picking up the second block, and correctly judging its orientation. The utility varies as a function of the duration of visual fixation on the placement bin. In this example, the movement durations have been constrained to equal the empirically observed movement times.

timized the timing of the saccade from the placement bin to the pickup location to maximize performance on the task. As stated previously, we defined maximizing performance as a combination of three factors: successfully placing the first block in the bin, accurately touching the second block to pick it up, and accurately judging the orientation of this block. The goal of our analysis is to determine the optimal timing of this eye movement, and then compare it with the empirically observed gaze behavior.

Figure 6 illustrates the calculation of the optimal eye movement behavior for one condition of the experiment (bin size, 8; aspect ratio, 1.05:1). In this figure, the three dotted lines plot the probability of successfully completing each component of the task (placing the first block, picking up the next block, and judging its orientation). The solid line gives the overall probability of success, defined simply as the product of these three curves. The figure shows that the expected utility strongly varies as a function of the timing of eye movements. This makes intuitive sense: if a subject fixates the placement bin for the entire trial, he or she never looks at the next block to be placed and is therefore at chance in determining its orientation. Performance in placing the first block asymptotes, since fixating the placement bin after the block has already been placed cannot improve motor performance in placing that block. If the subject only fixates the pickup location, information about block orientation is maximized, but accuracy in placing the first block will be at minimum since in this case the placement bin is far in the periphery for the entire movement. Probability of overall success is maximized at some intermediate trade-off between these two extremes. This point occurs when the combined utility curve (Fig. 6, solid line) reaches its peak.

The curves in Figure 6 were obtained by iteratively stepping through the entire range of possible fixation durations on the placement bin (using a step size of 10 ms) for a single experimental condition. The range of fixation durations examined was 0 ms, through the total duration of a round of the task (placing one block and picking up the second). At each level of fixation duration, we computed the probability of success for each of the task components illustrated in Figure 6. This procedure was repeated for all nine combinations of bin size and block aspect ratio. In

each case, the empirically observed motor durations were used as constraints on performance while the fixation durations were varied. The peak of the combined utility curve represents the predicted optimal duration of fixation on the placement bin. The results of this process are shown in Figure 7, for all nine experimental conditions. Each panel shows the utility curve for a different condition, with aspect ratio varying across columns, and bin size varying across rows. In each panel, the peak of the utility curve is marked by the dashed vertical line. The empirically observed fixation duration is indicated by the solid vertical line.

As can be seen by inspection of Figure 7, the optimal fixation duration closely corresponds to the empirically observed duration, in some conditions differing by <10 ms. The model predicts that the duration of fixation on the placement bin should vary as a function of both the bin size, as well as the aspect ratio of the block to be sorted. Since the effect of the aspect ratio occurs before the model fixates on the block, this effect reflects an anticipatory adaptation to the task: by deliberately shortening fixation while placing the first block, more time is reserved for the upcoming and more difficult perceptual judgment. These predicted effects are closely mirrored by the observed changes in fixation behavior.

It is notable that there is a relatively wide plateau at the peak of each utility curve in Figure 7, in which changes in the duration of fixation have only a small impact on predicted performance on the task. Despite this, humans appear to exhibit remarkable sensitivity in their behavior and control eye movements in a manner that is in close agreement with predictions from our model. The largest discrepancies between the model and empirical data occur in the largest bin size condition (radius, 24 mm). In this condition, subjects appear to fixate the placement bin longer than is strictly necessary to perform the task accurately. The model does not provide any strong insight as to the reason for this discrepancy, although the model does predict that this increase in fixation duration had little negative impact on performance in the task.

In predicting optimal fixation durations, our ideal actor model also implicitly predicts the optimal timing of the saccade away from the placement bin. The timing of this saccade can be examined relative to when the hand makes contact with the placement bin, a measure we have previously referred to as eye–hand delay. There are three reasons this analysis could be informative. First, given the small mismatches between human and optimal fixation durations, it is not obvious that the predicted eye–hand delay of the model would follow the same qualitative pattern of results as observed in our human subjects. Second, given that previous studies have found “look-ahead fixations” in complex tasks, one might expect that subjects would initiate a saccade to the second block before the hand has placed the first block. We did not observe this in our human experiment, and thus the question remains as to whether this reflects suboptimal

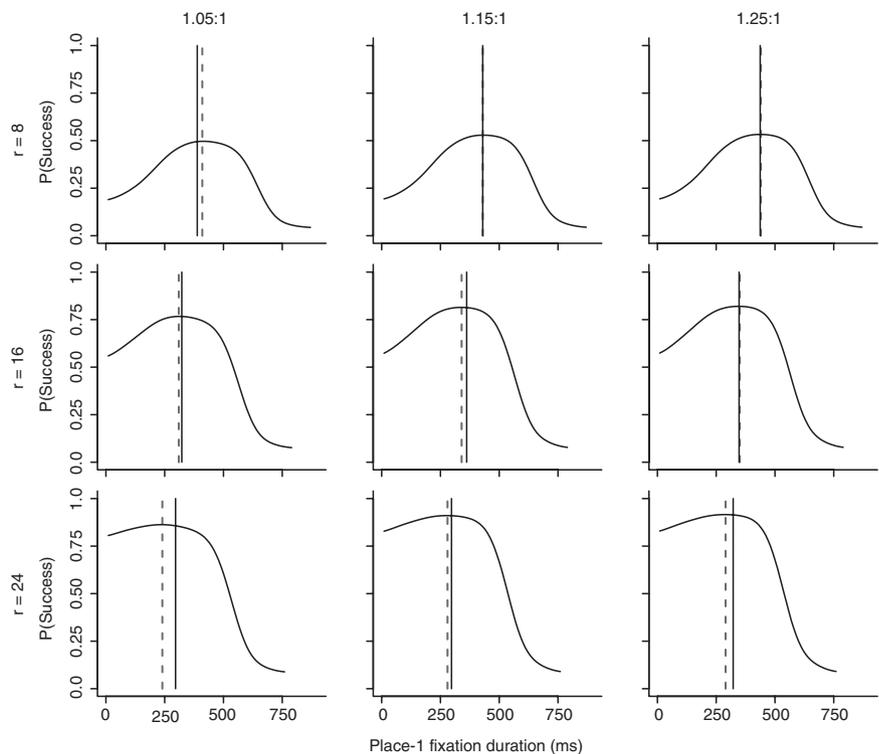


Figure 7. Predicted utility curves for each combination of block aspect ratio (columns) and bin size (rows). In each plot, the solid vertical line shows the empirical fixation duration on the placement bin; the dashed line indicates the point of maximum utility predicted by the model.

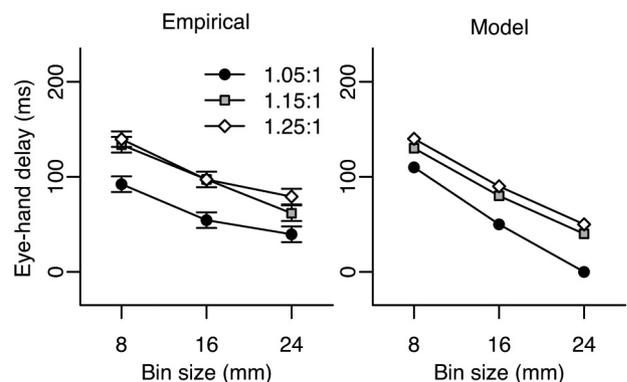


Figure 8. Empirically observed eye–hand delay (left) compared with the predicted optimal eye–hand delay (right). Empirical data are replotted from Figure 4*d*.

performance on the part of our subjects. Finally, an important claim of our research is that the coordination of hand and eye movements is not fixed but rather can flexibly adapt to changing task settings. If our model predicted a constant eye–hand delay, this would constitute evidence against the theory that our empirical results are the outcome of an adaptive coordination process.

Figure 8 compares the empirically observed eye–hand timing and the coordination pattern predicted by our ideal actor model. The figure shows that the optimal eye–hand coordination strategy is not fixed but rather varies with task conditions in a manner that closely corresponds to the observed data. Our model predicts that the optimal strategy is for the eye to remain fixated on the placement bin even after the finger makes contact. The magnitude of this eye–hand delay depends on both the radius of the placement bin and on the aspect ratio of the subsequent block.

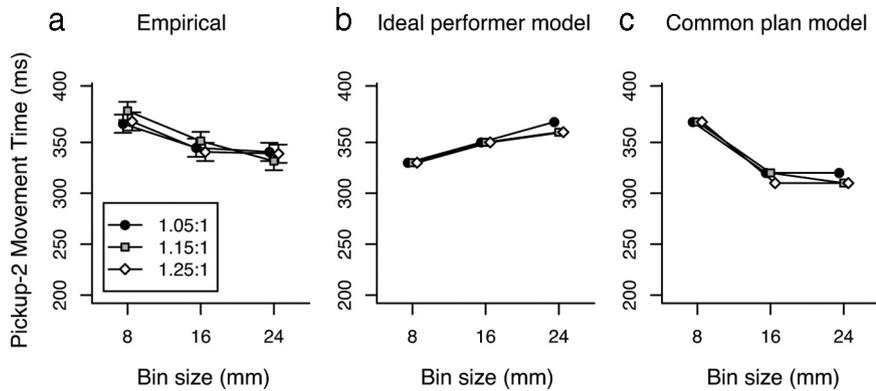


Figure 9. Movement duration to pick up the second block. *a*, Empirically observed movement duration. *b*, Predicted optimal movement duration. *c*, Model predictions revised under the assumption that a common motor plan (in terms of endpoint precision) was used for placing the first block and picking up the second.

Furthermore, the model predicts a comparatively large effect for the hard versus moderate aspect ratio conditions and a negligible difference between the moderate and easy aspect ratio.

Why does our model predict that the eye should linger on the placement bin after the finger has made contact? The answer to this question depends on the nature of the sensory noise in the visual system. When the hand is stationary, the visual system obtains more precise information about its location because of sensory noise proportional to hand velocity. Given the choice of making a saccade while the hand is moving, versus when the hand is stationary, the optimal solution is to make a saccade when the hand is moving, as less information is lost. The dependency of this effect on the bin size can also be explained by our model. When reaching to a larger target, the finger is more likely to make contact somewhere inside the placement bin even in the absence of visual feedback. Thus, at the end of the movement, the feedback control demands on the eye are decreased sooner, and the eye becomes free to move on to acquiring position and orientation information about the next block. This effect is also moderated by the aspect ratio condition of this subsequent block, with harder aspect ratios placing more demand on the visual system, and hence earlier saccades, than easier aspect ratios.

The model data shown in Figure 8 were obtained by fixing the movement durations to their empirically observed values for each experimental condition, and then optimizing the timing of eye movements with respect to the motor performance. A more stringent test of the predictions of the model is obtained by simultaneously maximizing performance with respect to both eye movement and motor movement durations. One difficulty in performing this comparison is that our model of the motor system does not incorporate variability in motor execution time; thus, the optimal strategy would be to adopt movement durations that took exactly the amount of time available in the task (2 s). Rather than building in temporal variability into our model, we chose to constrain the model to use the empirically observed duration of a complete round of the block-sorting task, using the empirical round duration for each experimental condition. We used this as a constraint on our ideal actor model but allowed it to distribute this time between the movement to place the first block and the movement to pick up the second. As before, motor and eye movement parameters were selected that maximized the overall performance of the model.

The predicted optimal duration of movement for picking up the second block is shown in Figure 9*b*, compared with the em-

pirically observed durations, replotted in Figure 9*a*. Recall that the unusual feature of the empirical data is that the movement duration to pick up the second block was found to increase with decreasing bin size, even though subjects were reaching to the same sized target in all conditions. The ideal actor model predicts the opposite pattern of results (Fig. 9*b*): return movements are shorter after placing a block in a smaller bin. The predictions of the model make sense for the task: if placing a block in a smaller bin takes more time to maintain sufficient accuracy, the model allocates more time to this motor task and consequently reduces the duration available to pick up the second block. In the next section, we discuss one possible alteration to our model that can account for

the empirical findings.

Finally, notice that there is a sizable effect of bin size, but no effect of the aspect ratio condition on the predictions of the model in Figure 9. Thus, aspect ratio had no effect on the optimal movement durations (Fig. 9*b*) but did have an effect on the optimal fixation durations (Fig. 8). Why should this be the case? The answer depends on the complex interplay of components in our ideal actor model, and thus there may be no single answer. However, one likely explanation is that varying the timing of eye movements had comparatively little impact on motor accuracy, compared with varying the duration of the motor movements. Given the choice of adapting movement durations versus fixation durations to accommodate the orientation discrimination task, varying the fixation duration had less impact on overall utility and is thus the optimal adaptation to the task demands.

Discussion

Situations with competing demands on vision are commonplace in routine human activities. For example, a person driving a vehicle may wish to reach to turn on the radio. Maximal motor accuracy on this task would suggest that the individual should fixate the control knob for the entire course of the movement. However, vision will likely be needed at the same time to watch for signs, plan upcoming turns, and for steering the vehicle. The challenge for the brain is allocating available resources to ongoing task demands in a manner that achieves goals efficiently.

In the study of eye movements, it has long been recognized that the goals of an individual strongly influence the resulting eye movements (Yarbus, 1967). This basic fact has been replicated in numerous domains, including driving (Land and Lee, 1994; Shinoda et al., 2001), sports (Land and McLeod, 2000), locomotion (Rothkopf et al., 2007), and so on (for recent review, see Land, 2006). Although these studies have focused on the effects of task on eye movements, few researchers have studied in detail the coordination of vision and motor control in complex or naturalistic tasks.

More recently, the development of lightweight eye trackers has begun to address this limitation. In one study of humans performing a routine activity (making tea) (Land et al., 1999), subjects frequently made a saccade to the next task-relevant object between 0 and 1 s before the previous motor action had been completed. These look-ahead fixations have been observed in other natural tasks as well (Pelz and Canosa, 2001; Pelz et al., 2001; Hayhoe et al., 2003; Mennie et al., 2007). This flexible de-

coupling of eye and hand movements is in contrast to the yoking hypothesis (Neggers and Bekkering, 2000) and instead suggests that the timing of eye movements is related to the need for sensory information at particular times and task locations. In a task in which gaze must be shared among multiple competing activities, once the demand for vision at the current fixation location is reduced, the eyes are free to move on to another component of the task.

Although these studies have demonstrated that visuomotor coordination is not likely explainable by reference to a task-invariant coupling of eye and hand, none of them has examined the issue of whether visuomotor coordination is optimal for a given task. We developed a task paradigm that allowed us to independently manipulate the demands on vision attributable to information acquisition and motor control. Given a richer task environment in which to observe behavior, we observed a corresponding increase in the richness of the behavior. This suggests that overly constrained task environments may unduly limit the complexity of behavior and thus limit theories of the adaptive control of cognitive, perceptual, and motor processes in interactive behavior.

Although our experiments provided evidence for an adaptive trade-off in visuomotor coordination, they did not provide evidence as to the utility of this trade-off. Was the pattern of behavior observed in our task optimal, or could subjects have performed better had they adopted a different strategy? To explore this possibility, we developed an ideal actor model for our task. Our model was used to predict the optimal visuomotor coordination strategy for the block-sorting task. Across the nine conditions of our experiment, the results from our analysis indicate that our human subjects exhibited behavior that was in close qualitative agreement with the predictions derived from our ideal actor model.

The one exception to this general finding is that our subjects exhibited correlations in their movement duration in placing one block and picking up the next. Although this behavior is not an optimal response to the task according to our model, it is worth considering possible alterations to the model that could account for the observed data. One hypothesis is that subjects planned both the block placement and the return movement in advance of placing the first block. When placing a block in a smaller bin, higher accuracy demands are necessary to maintain performance on the task, and thus a slower movement is adopted. According to our hypothesis, the movement duration for the return movement was programmed using the same endpoint accuracy demands as in placing the first block. That is, after placing a block in a small bin, subjects planned the return movement as if they were reaching to a smaller target.

This hypothesis was based, in part, on results in another experiment that demonstrated suboptimal motor planning in sequential pointing tasks (Wu et al., 2009). In the study by Wu et al. (2009), subjects executed a sequence of hand movements to touch two targets. When the reward associated with hitting the second target were increased, subjects were observed to slow down their movement to the first target—consistent with the idea that the accuracy demands for the second movement influenced motor planning for the first movement segment.

We implemented this assumption in a revised version of our ideal actor model, which we call the common motor plan model. In this revised model, the durations of motor movements were optimized under the assumption that the placement bin and the pickup area were the same size (i.e., using the same accuracy demands for both segments). The predicted movement durations under this revised model are shown in Figure 9c. With this addi-

tional assumption, the model is able to account for the empirically observed movement durations.

Although our revised model is able to account for the empirical data, we have no direct evidence to support the particular hypothesis underlying this model. It therefore remains possible that the observed “carryover” effects in our data are attributable to some other explanation. Since motor planning in sequential pointing tasks was not the direct focus of our experiment, exploring this issue further will be a question for future research.

Beyond demonstrating the adaptive nature of human performance in our task, our ideal actor model was able to account for relatively subtle features of behavior. In particular, the model was able to explain why we did not observe look-ahead fixations in our task, as have been observed in other studies (Land et al., 1999; Pelz and Canosa, 2001; Mennie et al., 2007). The optimal allocation of gaze in our task required subjects to maintain fixation on the placement bin even after the finger made contact to minimize visual uncertainty regarding the location of the hand. Importantly, this was not a fixed coupling of ocular and motor control, but rather the effect varied with both the bin size and aspect ratio manipulations in a manner quantitatively predicted by our ideal actor model.

Numerous previous studies have examined the question of optimality in visually guided motor behavior. These studies have focused on the optimal use of on-line visual feedback (Saunders and Knill, 2003, 2004), adaptation to costs on endpoint variability (Trommershäuser et al., 2006; Ma-Wyatt et al., 2010), or trading off target viewing time with motor execution time (Battaglia and Schrater, 2007). Our experiment differs from this previous work in several important ways. First, the optimal coordination of vision and motor control was not an obvious goal for our subjects; rather, subjects were merely instructed to rapidly sort objects based on visual features and were free to adopt any particular visuomotor strategy to accomplish this task. Second, executing our task required using the visual system to serve multiple roles: vision for on-line feedback control and using vision for acquiring information to subserve future motor acts. This functional division in the contributed roles of vision is akin to Ullman’s notion of visual routines (Ullman, 1984; Hayhoe, 2000). We believe this interleaved, sequential structure of behavior to be a ubiquitous feature of human performance in natural tasks, but rarely present in experimental studies of the visuomotor system.

In developing our ideal actor model for the task, we were forced to make several assumptions regarding the underlying mechanisms of visuomotor control. These assumptions included a hierarchically specified cost function on behavior and the assumption that, while fixating a block, humans are able to integrate sensory information about its position and orientation simultaneously. This latter assumption differs from assumptions made in another model of the sequential control of gaze (Ballard and Hayhoe, 2009), in which visual routines must compete for control of gaze. The extent to which visual routines exclusively control visual processing, and the extent to which multiple visual routines can simultaneously act on the same visual array, is an important issue in developing computational models of the control of gaze. The present work, however, has sought to underscore the importance of developing models that account for human interactive behavior as an adaptive process, rather than a set of fixed visuomotor couplings. Our convergent use of empirical study and optimality analysis provides substantial evidence in favor of the argument that routine human activity is characterized by the intricate and adaptive coordination of cognitive, perceptual, and motor processes.

Appendix: Details of the ideal actor model

Modeling the motor system

Following Liu and Todorov (2007), we model the hand as a point mass moving in a two-dimensional plane. The hand is controlled by forces that act in orthogonal directions; these forces are subject to low-pass filters that approximate the properties of human muscle. This simple model of the motor system can be described by a set of linear coupled differential equations. Let $\mathbf{p}(t)$, $\mathbf{v}(t)$, $\mathbf{a}(t)$, and $\mathbf{u}(t)$ represent the time-varying two-dimensional position, velocity, muscle filter state, and control signal applied to the hand, respectively. Then the continuous-time dynamics of the model motor system are given by the following:

$$\begin{aligned} d\mathbf{p}(t) &= \mathbf{v}(t)dt \\ m d\mathbf{v}(t) &= (\mathbf{a}(t) - b\mathbf{v}(t))dt \\ \gamma d\mathbf{a}(t) &= (\mathbf{u}(t) - \mathbf{a}(t))dt + \sum_{i=1}^c C_i \mathbf{u}(t)dw(t). \end{aligned} \quad (8)$$

The parameter m specifies the mass of the hand, b is a viscosity parameter that approximates the dampening properties of muscle, and γ is the time constant for the muscle filtering. These parameters were all set to the values specified by (Liu and Todorov, 2007). The vector $\mathbf{w}(t)$ indicates a continuous-time noise model defined by the Wiener process (this noise model is essentially the continuous-time version of discrete independent Gaussian noise). The term involving C_i scales the noise and represents control-dependent noise applied to the system, in this case given by the following:

$$C_1 = \begin{bmatrix} \sigma_{\parallel} & 0 \\ 0 & \sigma_{\perp} \end{bmatrix}, \quad C_2 = \begin{bmatrix} 0 & \sigma_{\perp} \\ -\sigma_{\perp} & 0 \end{bmatrix}, \quad (9)$$

where σ_{\parallel} and σ_{\perp} specify the SD of control-dependent noise added in the direction parallel and perpendicular to the applied force, respectively. The effect of this noise model is that large control signals result in higher variance in the forces applied to the hand, with larger noise occurring in the direction parallel to the direction of movement. The values of the noise parameters σ_{\parallel} and σ_{\perp} were chosen to fit the empirical motor noise observed in the experiment. In particular, we examined human performance while placing blocks in the right placement bin, in just one experimental condition (bin size, 8; aspect ratio, 1.05:1). Human motor variance in the principal and orthogonal directions of movement was computed by projecting the empirical error vectors onto the principal and orthogonal movement direction vectors. The principal direction of movement was defined as the vector connecting the pickup location to the center of the placement bin, and the orthogonal direction was the vector perpendicular to this. Principal and orthogonal motor variance was computed separately for each subject, and the values were averaged to obtain estimates of mean motor variance. Since the movement durations and gaze locations were recorded for our human subjects, we chose model parameters such that when a reaching movement was simulated using our model with the empirical movement duration, the resulting variance in the endpoint distribution was similar to the empirically observed variance.

Figure 10 shows the empirical (left) and model (right) motor endpoint variance for placing the first block. The figure shows the distribution of endpoints relative to the size of the placement bin (gray circle). The small black point indicates the mean placement location, whereas the ellipses represent 1 and 2 SDs of the endpoint distribution. For the model, the best-fitting parameters

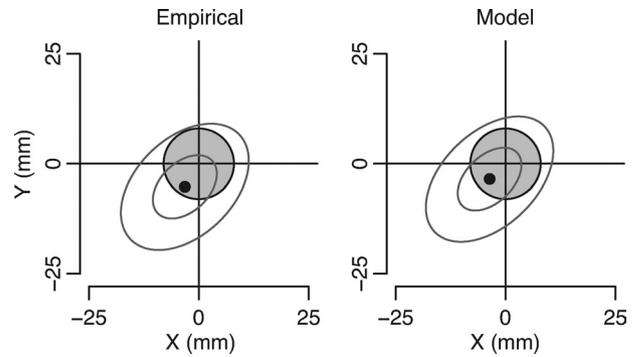


Figure 10. Comparison of empirical and model endpoint variance for placing the first block in the right bin, for the bin size of 8, aspect ratio of 1.05:1 condition. The gray circle represents the size of the placement bin, whereas the small black point shows the mean placement location. The ellipses represent 1 and 2 SDs of the endpoint distribution.

were found to be $\sigma_{\parallel} = 0.04$ and $\sigma_{\perp} = 0.023$, or in other words, the magnitude of noise in the direction parallel to movement was approximately twice the magnitude of noise acting in the perpendicular direction. Note that, for both the human and model data, there is an observed undershoot in the final position of the finger. In our model, this undershoot is governed by a cost parameter on the magnitude of control signals applied to the hand. Large control costs lead to increased undershoot, as the motor system trades off endpoint error with control costs. The quadratic cost function for the model is described in the main text.

In implementing the model, it is useful to combine all the variables that describe the state of the hand into the state vector $\mathbf{x}(t)$. For our simulations, it will prove convenient to also include the two-dimensional location of the pickup area (\mathbf{p}_0) and placement bins (\mathbf{p}_b) in the state vector, as well as a constant term. By including the pickup and placement bin locations in the state vector, it is possible to define the costs on motor performance purely as a function of the state variables. The resulting state of the system at time t is therefore given by the 11-dimensional vector $\mathbf{x}(t) = [\mathbf{p}(t), \mathbf{v}(t), \mathbf{a}(t), \mathbf{p}_0, \mathbf{p}_b, 1]$. These continuous-time state dynamics were discretized using the Euler approximation with a time step of 10 ms, leading to Equation 3.

In addition to examining motor endpoint error, it is also possible to examine the velocity profile of human and model reaching movements. Figure 11a illustrates 10 velocity profiles sampled from a single subject, in the bin size of 8, aspect ratio of 1.05:1 condition. The profiles were obtained by first smoothing the raw position data recorded from the OptoTrak system using a cubic smoothing spline, and then taking the first derivative of the position data to obtain velocity profiles. Since humans exhibited variability in movement duration from trial to trial, the profiles illustrated in Figure 11 are normalized in the following manner. The zero point on the abscissa corresponds to the time that the finger leaves contact with the work surface in picking up the first block in a trial, and the 100% point (shown by a vertical line) indicates the time of contact while placing that block. Each curve terminates at the time that the finger again leaves contact with the work surface after placing the block.

Figure 11b illustrates 10 simulated trajectories from the ideal performer model. For these simulations, the movement duration and contact durations were set to the mean values for the subject's data shown in Figure 11a. The visual fixation point of the model was set to the placement bin for the duration of the movement. Both human and model data exhibit approximately symmetric bell-shaped velocity profiles. Variability in velocity is

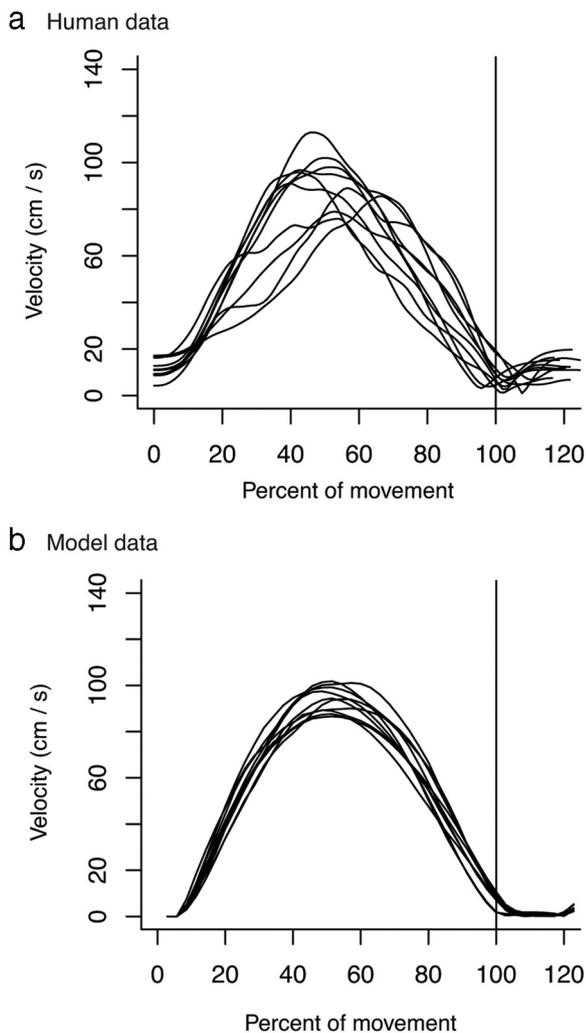


Figure 11. Comparison of empirical (top) and simulated (bottom) velocity profiles in reaching to place the first block on a trial. Human data are from the bin size of 8, aspect ratio of 1.05:1 condition. The vertical lines indicate the time of contact of the finger with the work surface when placing the block.

higher at the beginning of the movement for human subjects compared with the model. This is attributable to the fact that, at the start of each simulation, the state vector of the model was initialized to zero velocity and a constant location, whereas human subjects pick up the first block only after executing a reaching movement from the start cross. For the human data, variability in velocity appears to be higher throughout the movement compared with the model. However, the empirical data reflect both human motor variability as well as noise in recording the position of the hand using the OptoTrak system.

Modeling vision for on-line feedback control

On each time step, the model received noisy sensory information about both the position and velocity of the hand. Over time, these noisy signals are integrated by the model in an optimal manner using a Kalman filter (Kalman, 1960). The Kalman filter optimally combines incoming sensory signals with predictions based on the previous estimated state, and a forward model of the system dynamics (for other applications of a Kalman filter in modeling visual processing, see Saunders and Knill, 2004; Todorov, 2005).

Our sensory model included additive and multiplicative noise in both position and velocity signals. The matrix H in Equation 4 determines which state variables are observable; in our model, we assume that the perceptual system receives sensory information about the position and velocity of the hand, as well as the position of the target location that is currently being attended. Visual attention is assumed to be linked to gaze location, such that when subjects fixate the placement bin, they integrate sensory information about the location of the bin but not the pickup location. Gaze location is time-varying, so that when the eye saccades to the pickup area, the visual system begins integrating sensory information about this location.

We include two sources of multiplicative sensory noise in our model. First, sensory information about the position of the hand is corrupted by noise that is proportional to retinal eccentricity. We define the gaze location of the model at time t in workspace coordinates as $\mathbf{g}(t)$. The noise added to the visual location of an object with horizontal position \mathbf{p}_x is proportional to its eccentricity $|\mathbf{p}_x - \mathbf{g}_x(t)|$ and similarly for noise in perceiving vertical position. Psychophysical studies on two-point interval discrimination (Burbeck, 1987; Burbeck and Yap, 1990; Whitaker and Latham, 1997) have shown that uncertainty in estimating visual location can be closely described by a Weber fraction of 0.05. We therefore chose sensory noise parameters for our model such that when a static stimuli was presented in the periphery for 250 ms, the estimate by the model of its location was consistent this Weber fraction.

The second source of multiplicative noise in our model is sensory noise in the velocity signal for the hand. Psychophysical studies have shown that humans have a Weber fraction for motion discrimination of 0.08 (Mateeff et al., 2000). Furthermore, this value is mostly invariant to the retinal eccentricity of the motion. We chose sensory noise parameters such that the performance of the model in estimating the velocity of stimuli presented for 500 ms was consistent with the psychophysical results.

The sensory observation model defined by Equation 4 can be extended to incorporate time-delayed feedback as well. It is known that the sensory delay in the human visual system is approximately on the order of 100 ms (Wolpert et al., 2001). We include this in our model by maintaining a history of the 10 most recent states in the state vector $\mathbf{x}(t)$. The observation model (in particular the matrix H) extracts just the oldest state, so that the observer only has access to the time-delayed sensory signals. Since the step size in our discrete simulation is 10 ms, maintaining a history of 10 previous states results in an effective sensory delay of 100 ms.

Our model of the visual system also incorporates the effects of eye movements on available sensory information. When humans make saccadic eye movements, visual processing of object motion and location is mostly disrupted immediately before and during the course of the eye movement (Bremmer et al., 2009). We approximate these effects in our model by turning off incoming sensory information 50 ms before the execution of a saccade and during the course of the eye movement. Saccadic eye movements in our model were assumed to take 50 ms to execute. Thus, eye movements for our model resulted in a disruption of sensory information lasting for 100 ms. Finally, our model also incorporates variability in the timing of saccades: if a saccade was programmed to execute at time t , the actual time of the eye movement was drawn from a Gaussian distribution with mean t and a SD of 50 ms. As our simulation used discrete time steps, the sampled time was rounded to the nearest 10 ms.

References

- Abrams RA, Meyer DE, Kornblum S (1990) Eye-hand coordination: oculomotor control in rapid aimed limb movements. *J Exp Psychol Hum Percept Perform* 16:248–267.
- Ballard DH, Hayhoe MM (2009) Modelling the role of task in the control of gaze. *Vis Cogn* 17:1185–1204.
- Battaglia PW, Schrater PR (2007) Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *J Neurosci* 27:6984–6994.
- Becker W (1991) Saccades. In: *Vision and visual dysfunction, Vol 8, Eye movements* (Carpenter RHS, ed), pp 95–137. Boca Raton, FL: CRC.
- Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S (2010) The neural basis of the speed–accuracy tradeoff. *Trends Neurosci* 33:10–16.
- Bremmer F, Kubischik M, Hoffmann KP, Kregelberg B (2009) Neural dynamics of saccadic suppression. *J Neurosci* 29:12374–12383.
- Burbeck CA (1987) Position and spatial frequency in large-scale localization judgments. *Vision Res* 27:417–427.
- Burbeck CA, Yap YL (1990) Two mechanisms for localization? Evidence for separation-dependent and separation independent processing of position information. *Vision Res* 30:739–750.
- Carey DP (2000) Eye–hand coordination: eye to hand or hand to eye? *Curr Biol* 10:R416–R419.
- Chhabra M, Jacobs RA (2006) Near-optimal human adaptive control across different noise environments. *J Neurosci* 26:10883–10887.
- Diedrichsen J, Shadmehr R, Ivry RB (2010) The coordination of movement: optimal feedback control and beyond. *Trends Cogn Sci* 14:31–39.
- Fisk JD, Goodale MA (1985) The organization of eye and limb movements during unrestricted reaching to targets in contralateral and ipsilateral visual space. *Exp Brain Res* 60:159–178.
- Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol* 47:381–391.
- Gray WD, Sims CR, Fu WT, Schoelles MJ (2006) The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychol Rev* 113:461–482.
- Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. *Nature* 394:780–784.
- Hayhoe M (2000) Vision using routines: a functional account of vision. *Vis Cogn* 7:43–64.
- Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB (2003) Visual memory and motor planning in a natural task. *J Vis* 3:49–63.
- Heath M (2005) Role of limb and target vision in the online control of memory-guided reaches. *Motor Control* 9:281–311.
- Herman R, Herman R, Maulucci R (1981) Visually triggered eye-arm movements in man. *Exp Brain Res* 42:392–398.
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* 82:35–45.
- Keele SW, Posner MI (1968) Processing of visual feedback in rapid movements. *J Exp Psychol* 77:155–158.
- Kollmeier B, Gilkey RH, Sieben UK (1988) Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. *J Acoust Soc Am* 83:1852–1862.
- Körding KP, Wolpert DM (2004) The loss function of sensorimotor learning. *Proc Natl Acad Sci U S A* 101:9839–9842.
- 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.
- Land MF (2006) Eye movements and the control of actions in everyday life. *Prog Retin Eye Res* 25:296–324.
- Land MF, Lee DN (1994) Where we look when we steer. *Nature* 369:742–744.
- Land MF, McLeod P (2000) From eye movements to actions: how batsmen hit the ball. *Nat Neurosci* 3:1340–1345.
- Levitt H (1971) Transformed up–down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477.
- Liu D, Todorov E (2007) Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *J Neurosci* 27:9354–9368.
- Loftus GR, Masson MEJ (1994) Using confidence intervals in within-subject designs. *Psychon Bull Rev* 1:476–490.
- Mateeff S, Dimitrov G, Genova B, Likova L, Stefanova M, Hohnsbein J (2000) The discrimination of abrupt changes in speed and direction of visual motion. *Vision Res* 40:409–415.
- Ma-Wyatt A, Stritzke M, Trommershäuser J (2010) Eye–hand coordination while pointing rapidly under risk. *Exp Brain Res* 203:131–145.
- Mennie N, Hayhoe M, Sullivan B (2007) Look-ahead fixations: anticipatory eye movements in natural tasks. *Exp Brain Res* 179:427–442.
- Meyer DE, Abrams RA, Kornblum S, Wright CE, Smith JE (1988) Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychol Rev* 95:340–370.
- Myung IJ (2003) Tutorial on maximum likelihood estimation. *J Math Psychol* 47:90–100.
- Neggers SF, Bekkering H (2000) Ocular gaze is anchored to the target of an ongoing pointing movement. *J Neurophysiol* 83:639–651.
- Neggers SF, 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.
- Neggers SF, Bekkering H (2002) Coordinated control of eye and hand movements in dynamic reaching. *Hum Mov Sci* 21:349–376.
- Pélissier D, Prablanc C, Goodale MA, Jeannerod M (1986) Visual control of reaching movements without vision of the limb. II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Exp Brain Res* 62:303–311.
- Pelz JB, Canosa R (2001) Oculomotor behavior and perceptual strategies in complex tasks. *Vision Res* 41:3587–3596.
- Pelz J, Hayhoe M, Loeber R (2001) The coordination of eye, head, and hand movements in a natural task. *Exp Brain Res* 139:266–277.
- Prablanc C, Echallier JF, Komilis E, Jeannerod M (1979) Optimal response of eye and hand motor systems in pointing at a visual target. I. Spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information. *Biol Cybern* 35:113–124.
- Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. *Psychol Sci* 9:347–356.
- Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev* 111:333–367.
- Rothkopf CA, Ballard DH, Hayhoe MM (2007) Task and context determine where you look. *J Vis* 7:16.1–16.20.
- Sailer U, Eggert T, Ditterich J, Straube A (2000) Spatial and temporal aspects of eye-hand coordination across different tasks. *Exp Brain Res* 134:163–173.
- Saunders JA, Knill DC (2003) Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp Brain Res* 152:341–352.
- Saunders JA, Knill DC (2004) Visual feedback control of hand movements. *J Neurosci* 24:3223–3234.
- Shinoda H, Hayhoe MM, Shrivastava A (2001) What controls attention in natural environments? *Vision Res* 41:3535–3545.
- Todorov E (2004) Optimality principles in sensorimotor control. *Nat Neurosci* 7:907–915.
- Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Comput* 17:1084–1108.
- Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. *Nat Neurosci* 5:1226–1235.
- Trommershäuser J, Landy MS, Maloney LT (2006) Humans rapidly estimate expected gain in movement planning. *Psychol Sci* 17:981–988.
- Ullman S (1984) Visual routines. *Cognition* 18:97–159.
- Whitaker D, Latham K (1997) Disentangling the role of spatial scale, separation and eccentricity in Weber's law for position. *Vision Res* 37:515–524.
- Wolpert DM, Ghahramani Z, Flanagan JR (2001) Perspectives and problems in motor learning. *Trends Cogn Sci* 5:487–494.
- Wu SW, Dal Martello MF, Maloney LT (2009) Sub-optimal allocation of time in sequential movements. *PLoS ONE* 4:e8228.
- Yarbus AL (1967) *Eye movements and vision*. New York: Plenum.
- Zhang J, Bogacz R, Holmes P (2009) A comparison of bounded diffusion models for choice in time controlled tasks. *J Math Psychol* 53:231–241.