Visual Behavior and Planning for Object Manipulation:

Gaze Patterns for Altered Center of Mass

by

Michael Smith

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2017 by the
Graduate Supervisory Committee:

Marco Santello, Chair
Christopher Buneo
Sydney Schaefer

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

The interaction between visual fixations during planning and performance in a dexterous task was analyzed. An eye-tracking device was affixed to subjects during sequences of null (salient center of mass) and weighted (non salient center of mass) trials with unconstrained precision grasp. Subjects experienced both expected and unexpected perturbations, with the task of minimizing object roll. Unexpected perturbations were controlled by switching weights between trials, expected perturbations were controlled by asking subjects to rotate the object themselves. In all cases subjects were able to minimize the roll of the object within three trials. Eye fixations were correlated with object weight for the initial context and for known shifts in center of mass. In subsequent trials with unexpected weight shifts, subjects appeared to scan areas of interest from both contexts even after learning present orientation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

INTRODUCTION

It has been shown that subjects will fixate on perceived centers of mass when viewing a novel object and on the anticipated contact point of the index finger when reaching for precision grasp [1,2,9,10]. Additionally, eye fixation has been shown to precede and facilitate the dexterous manipulation of objects in many contexts during everyday activities [3, 7]. Visual fixations are tied to the specific anticipated tasks as well as the visual assessment of the objects, and are chosen based on the functional requirements [4,8,14]. Consequently, handling of an object specifically alters spatial representations and affects visual assessment of task relevant objects [11]. In studies where the lifting behavior of an object changes while the visual appearance is held constant, subjects show the ability to minimize the roll of an object within three trials for each new context [5]. In a related study, it was also found that explicit knowledge of the change in weight did not influence subjects' ability to counter the roll of a previously explored object [6,7]. Subjects react to new changes in an object's center of mass with the same latency regardless of explicit knowledge about the location of weight – similar to encountering the object for the first time. When switching between similar tasks of opposing context, sensorimotor memories from a previous task can impede the learning of a new paradigm, but previously learned tasks can be more successfully returned to if mastered before the new context [12,13]. These processes are respectively known as interference and retention. Currently, no investigation of the visual assessment of objects during interference and retention has been performed. This study investigates the connection between eye fixations and objects that are visually symmetrical but mechanically asymmetrical. Changes in center of mass that are both known and unknown to subjects are explored in relation to minimizing roll and patterns of eye fixations. We hypothesized that (1) changing center of mass would modify subjects' preferred visual fixation points as they learned the new task and (2) conscious changes in object orientation would encourage subjects to visually reassess the object as novel.

MATERIALS AND METHODS

Subjects

Ten right-handed subjects (five male, five female, median age 20) participated in this experiment. All subjects were informed of the physical requirements of the task and gave written consent according to IRB approval. Subjects were divided evenly into two groups based on the order of perturbation directions. The first group experienced perturbations to the left and then rotated the object to the opposing context. The second group experienced perturbations to the right first and then rotated to the opposing context. Both groups followed the same protocol and grouping of trials.

Setup

Subjects were seated in a chair inside of a metal frame (75" x 39" x 35.5") (*figure 1 a-b*) with a chin rest placed at head level. They were allowed to adjust the frame and chair until a comfortable position was found.

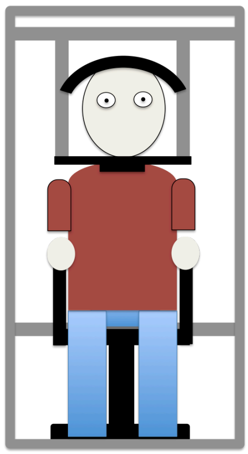

figure 1a: front view of subject in frame

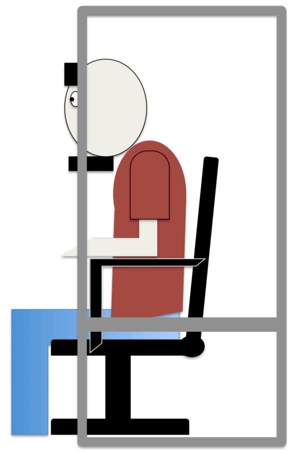

figure 1b: side view of subject in frame

A desk was placed in front of them with a hand sensor and an inverted t-shaped object (*figure 2*). The object is visually symmetrical, with compartments in the bottom (left, middle, and right) that allowed for the weight distribution to be made uneven without changing the total weight of the object. Subjects were free to grasp the object anywhere along the length of the handle during trials, so that object manipulation would be accomplished by altering digit center of pressure [9]. A 400g weight was placed in one of the three compartments depending on the block of trials. On either side of the object is an ATI Nano-17 F/T sensor attached to a flat plate that allowed for a free range of finger placement along the handle. A 3-axis Polhemus motion-tracking sensor was affixed to the top of the object.

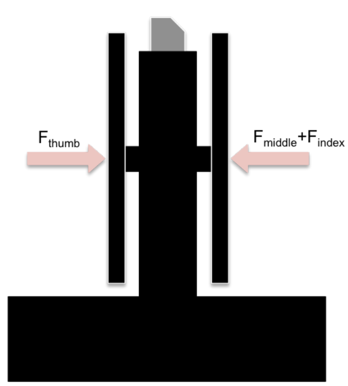

figure 2: the inverted T-shaped object. Compartments in the bottom allow for uneven weight distribution while maintaining visual symmetry. Subjects are free to place fingers anywhere along the handle.

Subjects' eye movements were monitored using an SMI ETG Eye Tracking system. The system was held in place on their head throughout the experiment. Data recording was synced between force sensors, polhemus, and eyeview through the use of a cortech trigger system.

Experiment

Subjects performed 118 trials over 5 blocks. In between trials they were instructed to keep their eyes closed, to confine any visual assessments to the experimental recordings. Subjects listened for verbal cues from a labVIEW program, 4 in total. The first cue was to open their eyes, the second was to lift the object (5 seconds later) the third was to lower the object (3 seconds after crossing a height threshold) and the last was to close their eyes (another 5 seconds). Labview data was recorded for a total of 17 seconds for every trial, but trials could be shorter than this time frame depending on subjects' speed. Prior to lifting subjects were instructed to keep the object parallel to the ground and to lift using only index and thumb of the right hand, placed on opposite sides of the object handle. Subjects were given periodic reminders to maintain object orientation.

The first block that all subjects encountered was a series of 20 "null" trials, in which a weight was placed in the center compartment of the object. For the next block of 15 trials, the weight was switched to either the left or the right side, depending on group number. The weight applied a torque of 255 N*m in the respective direction during lift, which subjects were required to counter. In the third block, subjects repeated sequences of one "rotate" trials and three "initial" trials twelve times (48 trials total). In rotation trials, subjects were asked to rotate the weighted object 180º without lifting it from the table – placing the weight on the opposing side. After one trial they returned it to the initial context and performed three more lifts. Following this block, subjects were given a short break and then completed another block of 20 "null" trials and another block of 15 trials with the initial weight context. Subjects were not told which direction the weight would be pulling. A visual layout of this trial paradigm can be found in figure 3

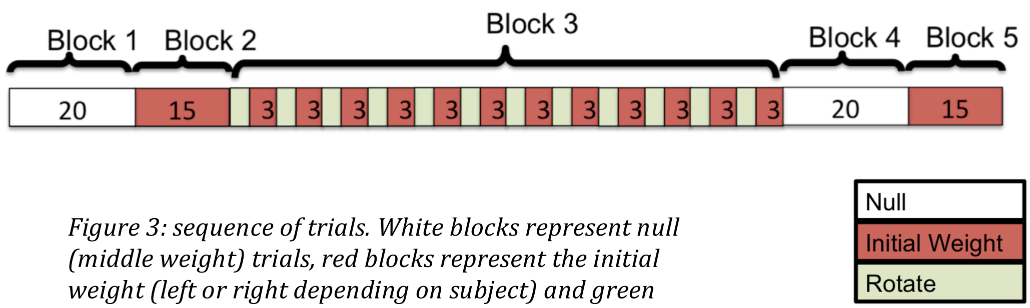

Figure 3: sequence of trials. White blocks represent null (middle weight) trials, red blocks represent the initial weight (left or right depending on subject) and green blocks represent rotation trials (opposite context).

Data Analysis

Data from all systems was synchronized using cues from a Cortech 24-bit Trigger I/O interface. Unique signals were sent for each of the 4 cues that subjects received during trials. Values for reaction time (T_r), Peak Roll (*PR*), and Time of Peak Roll (T_{pr}) were computed for all trials of all subjects to quantify movement behavior. Roll was calculated as the root mean square of rotational movements in the x, y, and z direction recorded by the Polhemus during a trial. The maximum value reached by roll following the go cue was taken as *PR*. Reaction time was calculated as the time between the "go" cue and the onset of lift, as determined by movement in the Polhemus and the load force on the object. Time of Peak Roll was found by subtracting the lift onset from the time at which peak roll occurred. Comparisons were made between 6 blocks of trials: two null, three with initial weight, and one rotation. Rotation trials were pulled from the 3rd block and plotted as a separate condition, labeled "Block R". Residual analysis was performed on all behavioral variables to check for the presence of heteroskedasticity, and a log transformation was performed on peak roll to remove non-constant error variance. Comparisons were made between all blocks using 1 way ANOVA with repeated measures on the first three trials of each. Responses were analyzed with "treatment" as a factor of interest and subject as a blocking variable. Tukey pairwise comparisons were performed to find differences between blocks and t tests were used to find differences within blocks.

Times and locations of eye fixations were extracted from the eye tracking data. For each trial, the location of the object in the field of view was recorded and used to create a 10 x 10 grid of possible fixation locations (*figure 4)*. The top and bottom of the left side of the handle were used to anchor the grid for each instance, and size was scaled relative to object's size in frame.

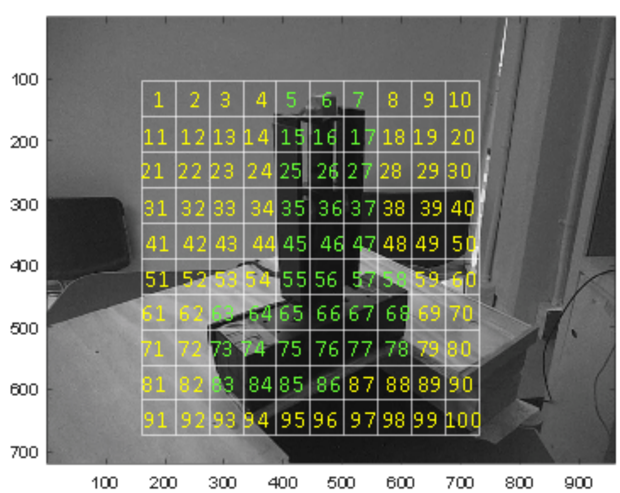

figure 4: a view of the object from one trial, with overlaid grid. Green numbers represent fixations that are on the object. X and Y axis are in arbitrary units relative to screen

Fixations from the planning phase of the task were recorded at the value of the box they fell within. To account for potential foveal and parafoveal components of vision, a convolution was performed on fixation counts. A fixation within a box was scored as 0.2 for the direct location and 0.1 for all bordering locations. Convolution maintains the validity of regions of interest on the object by using a weighting scheme to compare and distinguish adjacent fixations. The aim of this process was to reduce variability caused by relatively high degrees of freedom. Heat maps were generated for convoluted values of each trial and average across the first and last three trials for each condition. A heat map for the first and last three trials from each condition was then computed by averaging across all subjects. For subjects that started with a right-weighted object, fixation maps were mirrored about the Y-axis to pool data for the two conditions.

To investigate significant differences in gaze maps, a random sample bootstrapping method was used to test hypothesis on two different statistics of interest. The first of these was correlations between 2d grids of fixation counts, where each grid contained the average number of fixations across three learning trials for each subject. A single 10x10 grid of average fixation count for all subjects (30 trials total) from one block was correlated with a second set of trials from a different block, and this correlation coefficient was compared with a series of 1000 coefficients from randomly sampled grids to determine potential significance at P=0.05. The same process was used for comparisons of subtractions between grids. Grids composed of differences between two conditions of interest were compared with randomly sampled grids of differences to determine which values fell within a significant tail of the generated normal distribution, indicating that the difference was non-zero.

The number of fixations per trial was also analyzed for more quantitative metrics. Number of unique fixations was correlated with number of total fixations. Comparisons were made from the beginning of trials to the end of trials using a 2 way ANOVA to examine effects from blocks and subjects on eye fixations.

RESULTS

Reduction of roll

Figure 5 shows a time course of object rolls for all trials from a representative subject, with first and last trial highlighted in each block. Figure 6 shows average peak rotations across all subjects and all trials, separated by color.

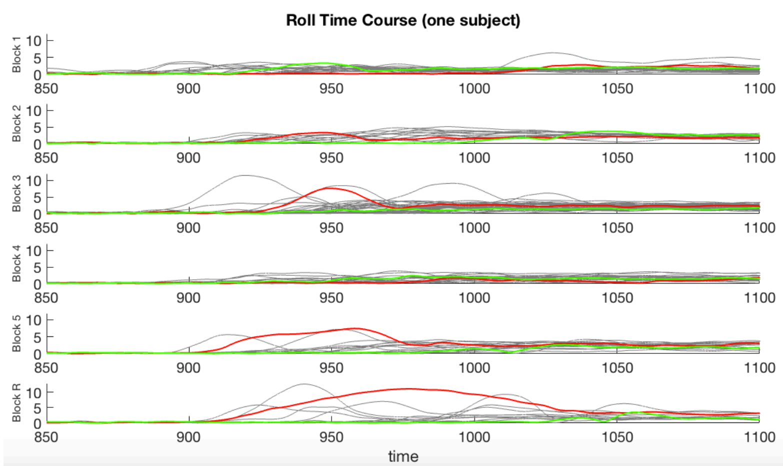

Figure 5: Time course plots of all roll values of all trials for subject 5 between 850 and 1100 ms (rough window of lift). Red line represents the first trial in a block and green represents the last. Block R is composed of the 12 rotation trials that are interspersed evenly within block 3.

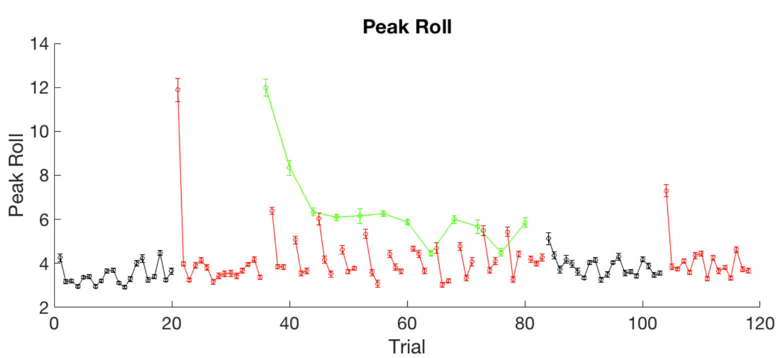

Figure 6: Average peak roll values for all trials. Null blocks (1 & 4) are shown in black, initial weight blocks (2, 3, & 5) are shown in red, and rotation block (R) is shown in green.

In all conditions except for the first (null) block, subjects showed a significant reduction ($P < 0.05$) in PR from the first to the last trial. Figure 7a shows average *PR* for first and last trials of every block. Initial trial peak roll (Figure 7b) was lowest for the two null blocks (1 and 4) and highest for the two "first encounter" blocks (2 and R) where subjects initially experienced each rotation context. In all conditions except for the rotation trials (Block R) subjects reduced *PR* to similar levels by the final trial (Figure 7c).

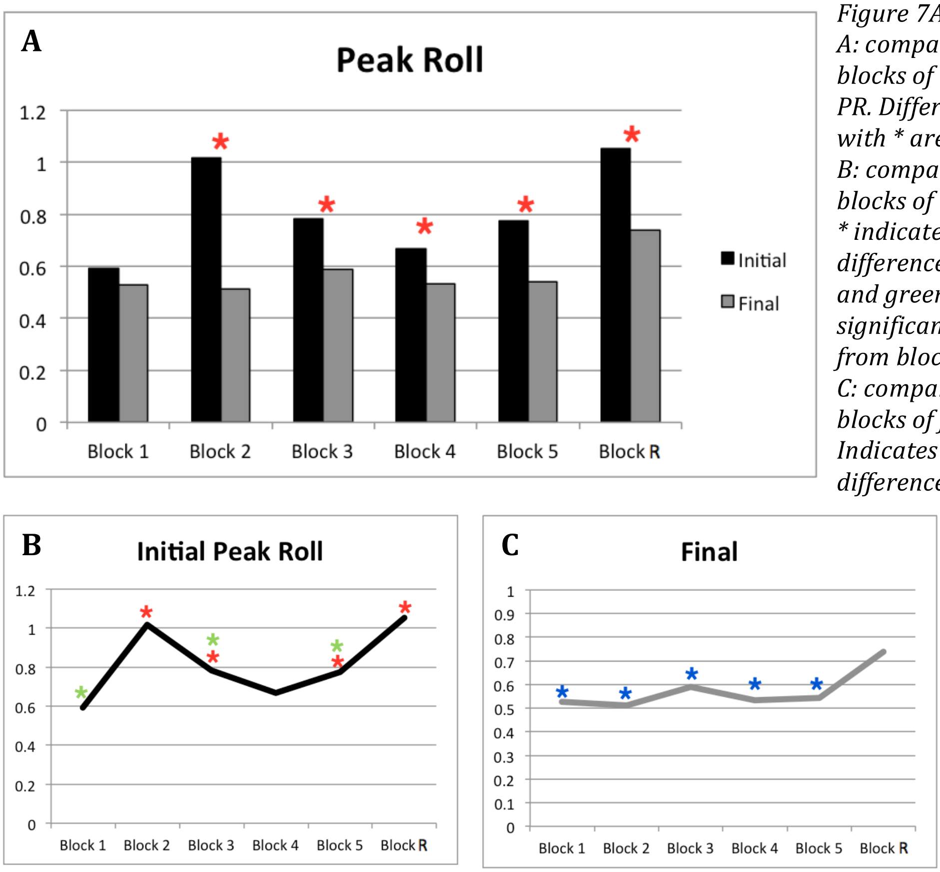

Figure 7A-C
*A: comparison within blocks of initial and final PR. Differences marked with * are significant*
*B: comparison between blocks of initial PR. Red * indicates significant difference from block 1 and green * indicates significant difference from blocks 2 and R*
*C: comparison between blocks of final PR. * Indicates significant difference from Block R*

From ANOVA comparisons, significant interactions were found from both subject and block ($P < 0.001$ and $P < 0.01$ respectively). This indicates that despite high subject variance there is also a large difference as a result of the learning conditions for peak roll. Log values were used for comparisons in peak roll to account for heteroskedasticity in response values (figure 8).

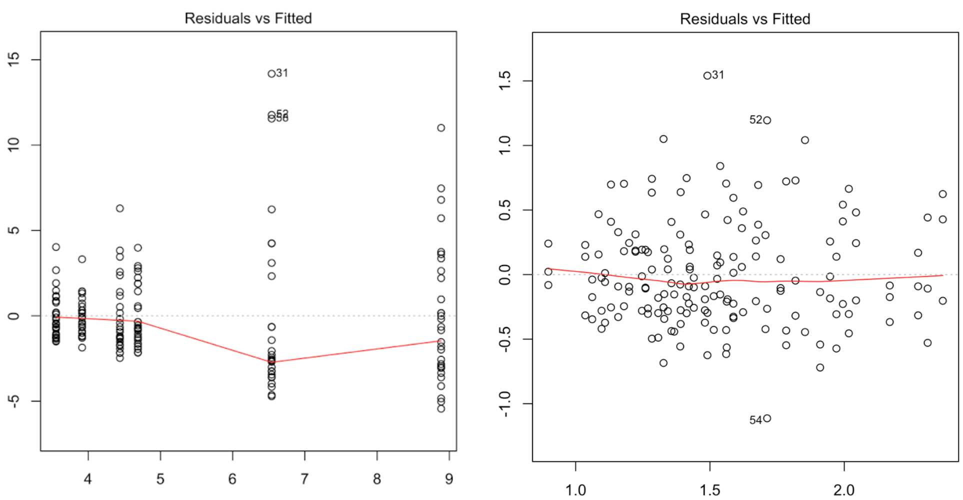

Figure 8: Residual plots for standard (left) and log transformed data from peak rolls. Increased variance with higher response levels shows nonconstant error variance. The log-transformed data shows that variance is more constant as a function of response level

Timing effects

Fewer significant results were seen in the timing of peak roll. In block 3, peak roll occurred significantly sooner in the first trial than it did in the last (figure 9A). Between blocks, 1 and 3 presented significantly different initial values for T_{PR} (figure 9B). For final T_{PR} values (figure 9C), significant differences were found from the first to the second, third, and fourth blocks and from the third to the fourth and fifth blocks. No significant effects were found on values of Reaction time for any block.

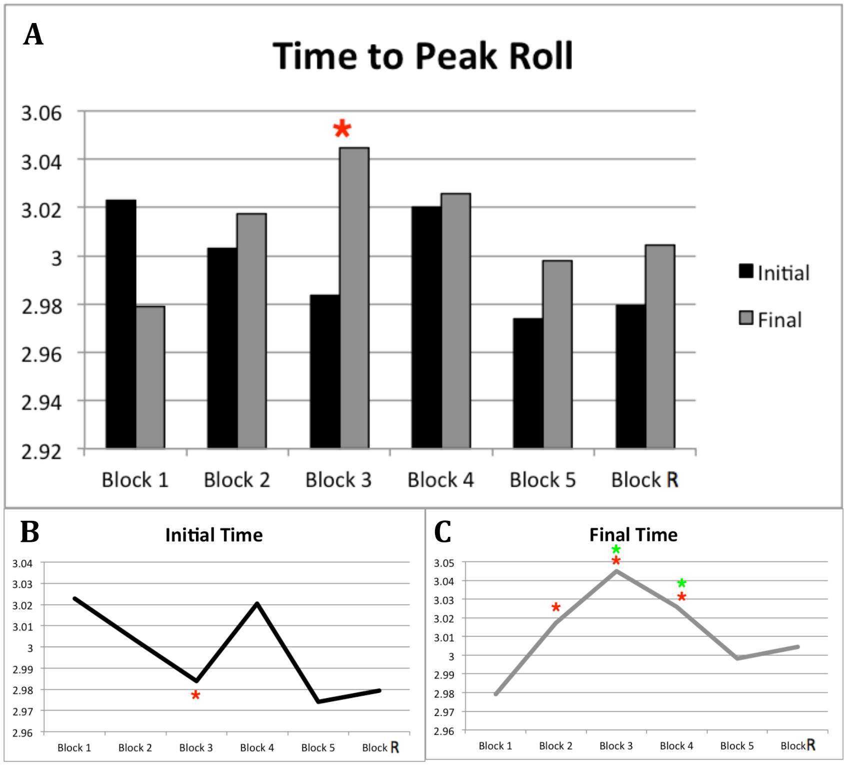

Figure 9A-C
*A: comparison within blocks of initial and final T_{PR}. Differences marked with * are significant (P<0.05)*
*B: comparison between blocks of initial T_{PR}. Red * indicates significant difference from block 1 (P<0.05)*
*C: comparison between blocks of final T_{PR}. * Indicates significant difference from block 1 and green * indicates significant difference from block 5 (P<0.05)*

Rotation Block

Comparisons for the first and last trials of the rotation sequence were made to verify reduction of roll while learning two contexts. Comparisons were made using trials pre and post the first and last rotations (Block 3 trial 1 vs. Block R trial 1 and Block 3 trial 36 vs. block R trial 12). PR spiked significantly after the first rotation and was relatively equal after the last rotation (figure 10).

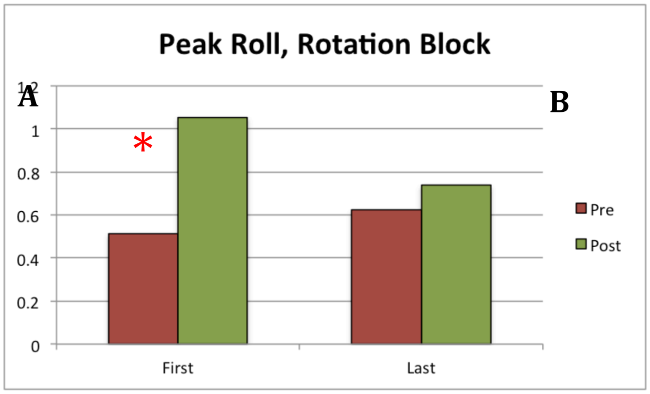

Figure 10: Comparisons for magnitude of Peak Roll in the trial immediately before and after subject rotates the object. Asterisks indicate significant differences (P<0.05)

Eye Fixations

No notable differences were found in the number of fixations performed during the first and last trials of each block or between values of fixations for all blocks. A series of averaged heatmaps for the beginning and end of all conditions can be found in figure 11A:L. These serve as broad representative maps of where subjects were looking as they progressed through the series of rotations and weight shifts. Using a bootstrap comparison with an accepted significance value of $P = 0.1$, subtractions between conditions were assessed. Figure 12A:E shows comparisons between the first three of every block with the last 3 of block 1. These comparisons represent the difference between gaze for learning a new context (or repeating with interference) and gaze for a context that has been successfully learned. In each plot, the initial set is subtracted from the second. Squares with lower color values (more blue) indicate that the number of fixations was lower in the second set of heatmaps, and higher values (more red) indicate that the number of fixations was higher in the second set. Values that are considered statistically significant ($P < 0.1$) from a randomly generated model are marked with diagonal stripes.

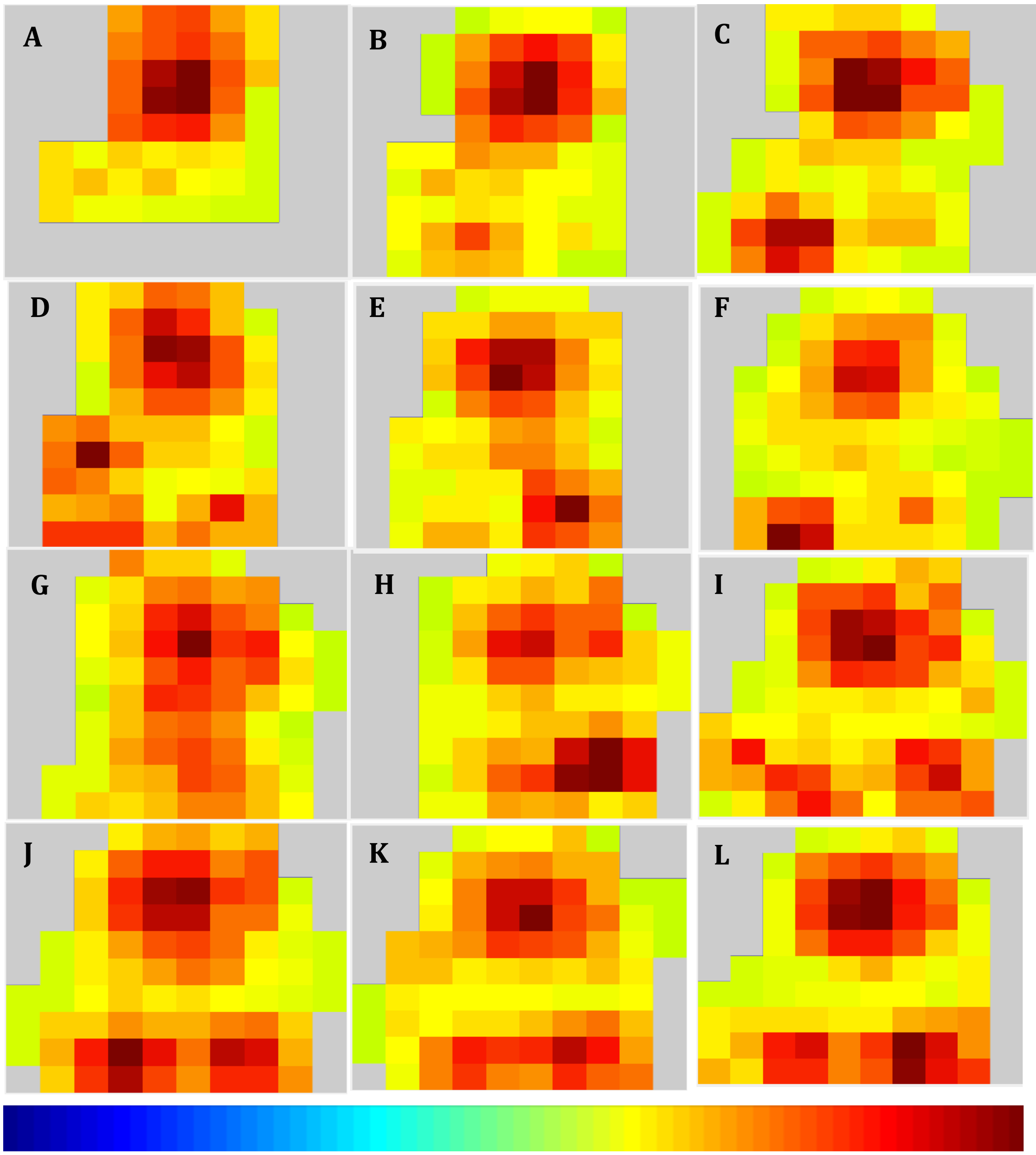

figure 11A-L : composite heatmaps showing averaged eye fixations for the first and last three trials of every condition as follows: A&B: Block 1 (null), C&D: Block 2 (initial weight), E&F: Block 3 (initial weight between rotations), G&H: Block R (rotations in between block 3), I&J: Block 4 (return to null), K&L: Block 5 (return to initial weight). Values are assigned based on number of distinct fixations. Scale bar is on the bottom, where green (middle) represents 0 value. Grey areas represent no fixations

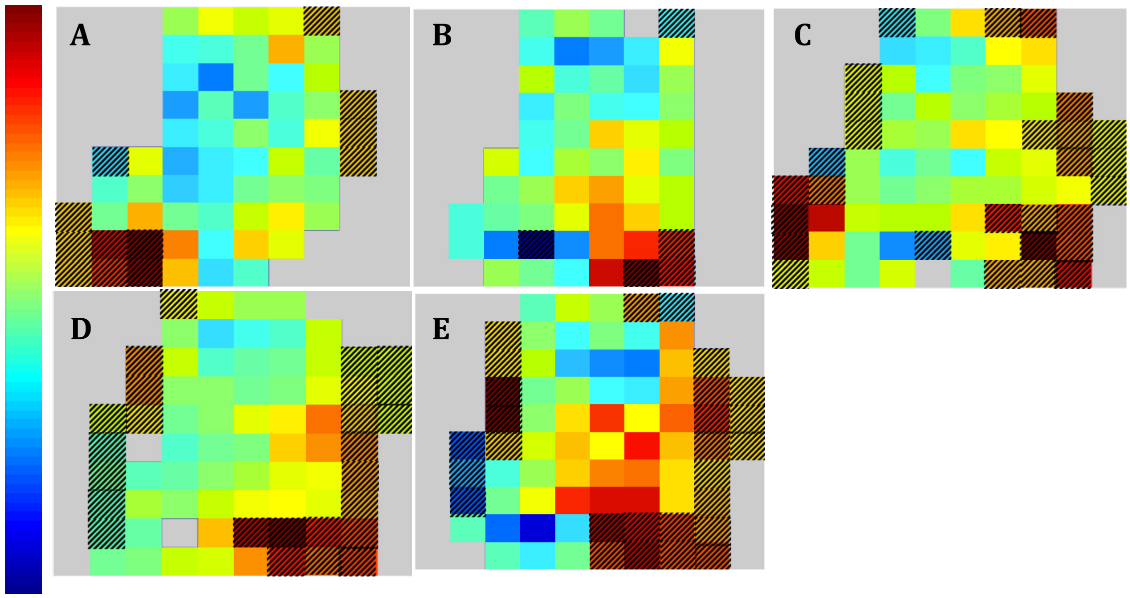

figure 12 A:E – comparisons between the last three trials of the initial null context and the first three of each successive context. The last three trials of the null represent gaze patterns when a subject is comfortable with object manipulation, and the first three of the others represent gaze patterns when a subject is learning a new context. Comparisons are as follows: A: second block, B: third block, C: fourth block, D: fifth block, E: sixth block (rotation trials). Significant differences (P < 0.1) are marked with diagonal lines.

Rotation Block

Similar to the comparisons made with peak roll, the pre- and post- gaze patterns were compared for the first and last rotations of the joint block 3 and R set. The representative maps and the subtractions are shown with significance in figure 13A:B. A large amount of differences were found accounting for significant changes as well as statistically significant consistency. To examine any effects that the rotation trials had on the recall of null and initial context manipulations, a final comparison was made between the averaged heatmap of learned rotations (figure 11H) and the heatmaps of block 4 and block 5 (figure 11J & L). The results of this comparison can be found in figure 14A:B.

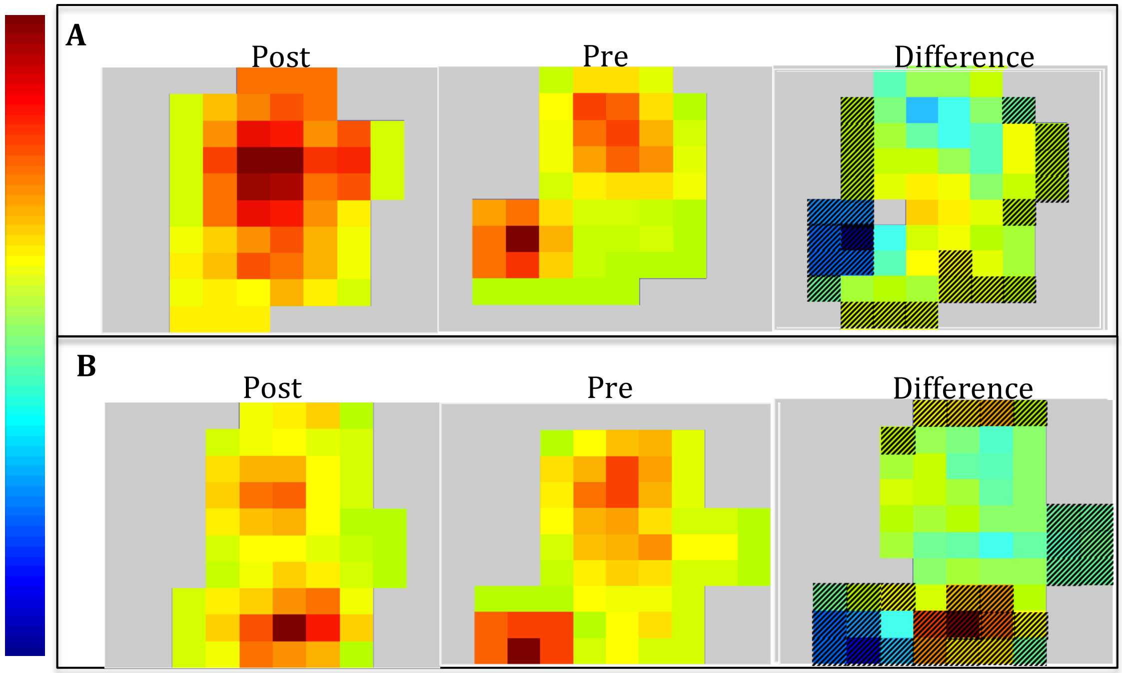

figure 13A-B: Pre and post rotation gaze maps for the first and last of the 12 rotations. Pre is subtracted from post to produce difference maps, and significant differences are marked with diagonal lines. Areas that are more blue indicate higher fixations before the rotation, and areas that are more red indicate fewer fixations before the rotation

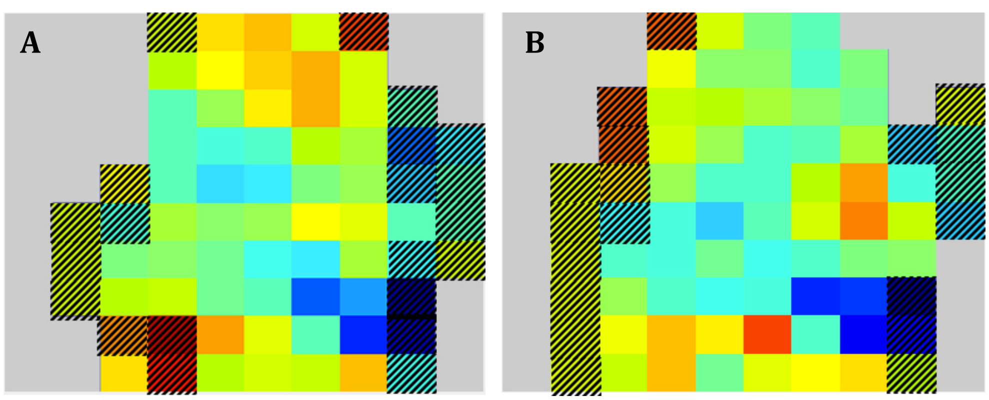

Figure 14A-B: subtraction comparisons between the gaze maps of the last three rotations trials with the last three trials of block 4 (A) and the last three trials of block 5 (B). These maps represent the effect that rotation has on gazemaps for recalling null and initial context trials.

DISCUSSION

Reduction of Roll

As expected, subjects performed the 20 null trials of block 1 with very little error in the form of roll. Treating this as a baseline for task performance, each successive switch of context induced significant error in all subjects, as evidenced by the increase in PR. However, after each block subjects showed reductions in PR that indicated that object manipulation was being learned for the new weighting paradigm. In all block except for the single rotation trials (block R), subjects returned to the same level of PR as they initially started with. The reduction that did occur in block R suggests that subjects would be capable of learning to reduce the roll in these rotations with a longer sequence of trials. Pre- and post- rotation comparisons support this by highlighting the difference in PR before and after the subject has completed the sequence. The first rotation induces a large spike in error, while the last one shows no significant difference. Despite initial difficulties, subjects were able to reduce the roll of the object while practicing single trials of a new context in between three trials of a previously learned one. This is seen most clearly in figure 6, where the green lines represent rotation trials interspersed with red initial context trials. This reduction in roll across block 3 indicates that subjects are initially not as proactive at correcting for the roll, despite being aware of which side the weight is on.

Overall, these results concur with the findings of Zhang *et. al. (2010)* – subjects do not immediately modulate behavior to anticipate roll when rotating an object, but they are capable of learning over the course of several individual trials.

Patterns in Eye Fixations

Composite heatmaps of eye fixations seem to indicate that the current and previously encountered center of mass of the object directly affects where the subject is looking, especially during initial learning of a context. Throughout the first null block, concentration of fixations is generally around the center of the handle, where subjects place their fingers during the lifting phase. This is true of the averages of both the first and the last three trials. In the three trials where subjects are learning the initial context, gaze shows two hotspots: one on the center of the handle and one on the side where the weight is placed (left side of figures due to mirroring of right-sided trials). Bootstrapping analysis supports this claim, as there is a concentration of statistically significant increases in the bottom left of the first subtraction map (figure 12A). This spot appears to grow in intensity as subjects complete further trials in the same context. In the three trials following the first rotation trial, this focus appears to shift to that of the opposite context, as evidenced by the slight concentration of fixations on the bottom right of the second comparison (figure 12B). This seems to be related to the gaze patterns of the first three rotations, in which a highly significant number of fixations are focused on the side where subjects are anticipating the weight.

Interestingly, when subjects return to the null condition, their gaze patterns do not resemble the initial encounter. Fixations generally cover a larger distance, with significant increases on both sides of the weighted portion. Despite not needing to anticipate a perturbation, subjects continue to scan the potential areas of the object that could lead to object roll. This bias persisted into the following set of recall trials, in which subjects experienced the initial weight context for a final block. There continues to be a significant concentration of fixations towards the side of the object where there is no weight.

Eye Fixations and Rotations

Analyzing the trials pre- and post- rotation again highlight the subjects' ability to internalize two contexts simultaneously. In the first comparison, rotation causes a sharp decrease in fixations on the initial side – subjects appear to return focus to the center of the object, as if reassessing it as a new object. Significant differences appear as decreases in focus on the previously weighted side of the object and a generally increased spread of gaze around the object. After the last rotation, it becomes apparent that subjects are using distinct gaze patterns for each context – subtracting the "pre-" from the "post-" shows a significant decrease in fixations on the initial context side and an increase towards the flipped context side.

Gaze pattern appears to be tied to the weight of the object most directly for trials where subjects perform the rotation themselves. After completing the rotation block, subjects made clear changes to gaze pattern that indicated an awareness of the object's condition. However, when the weight was shifted for them, they did not adjust their gaze patterns as distinctly – even when they had returned to a baseline of *PR*.

CONCLUSION

Results from this experiment indicate a conscious change in fixation patterns when subjects manipulate the context of an object themselves, but not one that carries over when the context is switched for them. Despite reducing the peak roll in every condition, subjects only show specific and context relevant fixations during the first encounter with the weight and the successful rotation trials. In subsequent encounters with the initial context and while learning rotation trials, subjects scan the object more generally, with fixations concentrated on both sides of the object as well as the center. Although the gaze patterns in the planning phase are correlated with subject's anticipation of roll, they do not seem critical for the reduction to baseline levels. Subjects are capable of completing the task without specific fixations on the area of interest. The implication is that eye fixations may be tied to subjects' previous interactions with the objects, rather than their anticipated actions during the lifting phase.

Further work can help to strengthen the results of this study. In particular, the current set of 10 subjects is potentially too low to fully explain the patterns in gaze, and is noted as the reason for a significance value of $P = 0.10$. The current grid setup specifies 100 possible points of fixation, which discretizes the essentially limitless number of fixations that a subject could make on the object. At the extreme limit, each of the 684,000 pixels on the screen could function as a recognizable fixation point for a subject to land on. With subjects averaging 8 ± 3 fixations per trial, the grid is required to observe commonalities in eye location. By increasing the number of subjects in the study, more specific conclusions can be drawn about locations of eye fixations. Additionally, the object in this study is skewed relative to the subject. While this did not prevent subjects from fixating on the further side, an experimental setup that does not require this angle would further strengthen results. This was discussed during the planning phase, however many cases that allow for this require hand angles that are potentially uncomfortable for long periods of time.

REFERENCES

[1] Juravle, G., *et. al.* (2015). The hand grasps the center, while the eyes saccade to the top of novel objects. Frontiers in Psychlogy, 6:633

[2] Brouwer, A., *et. al.* (2009). Differences in fixations between grasping and viewing objects. Journal of Vision, 9(1):18, 1-24

[3] Land, M., *et. al.* (1999). The roles of vision and eye movements in the control of activities of daily living. Perception, 28, 1311-1328

[4] Belardinelli, A., *et. al.* (2016). It's in the eyes: planning precise manual actions before execution. Journal of Vision, 10.1167/16.1.18

[5] Fu, Q. *et. al.* (2010). Anticipatory planning and control of grasp positions and forces for dexterous two-digit manipulation. J. Neurosci, 30(27): 9117-9126

[6] Zhang, W., *et. al.* (2010). Manipulation after object rotation reveals independent memory representations of digit positions and forces. J Neurophysiol, 103: 2953-2964

[7] Fu, Q., *et. al.* (2014). Learned manipulation at unconstrained contacts does not transfer across hands. PLOS, 10.1371.0108222

[8] Johannsson, R.S., *et. al.* (2001). Eye-hand coordination in object manipulation. The Journal of Neuroscience, 21(17): 6917-6932

[9] Desanghere, L, and Marotta, J. (2015). The influence of object shape and center of mass on grasp and gaze. Frontiers in Psychology 6:1537

[10] Zhou, X., *et. al.* (2005). Center of mass attracts attention, Neuroreport, 17(1): 85-88

[11] Draschkow, D. and Vo, ML. (2016). Of "what" and "where" in a natural search task: active object handling supports object location memory beyond the object's identity. Atten Percept Pschophys, 78(6):1574-84

[12] Fu Q, Santello M (2012). Context-dependent learning interferes with visuomotor transformations for manipulation planning. J Neurosci 32:15086-15092

[13] Fu Q, Santello M (2015). Retention and interference of learned dexterous manipulation: interaction between multiple sensorimotor processes. J Neurophysiology 113:144-155

[14] Kjellstrom, H (2008). Simultaneous visual recognition of manipulation actions and manipulated objects. Computational Vision and Active Perception Lab.