The role of gaze and prior knowledge on allocentric coding of reach targets

Zijian Lu Department of Experimental Psychology, Justus-Liebig-University, Giessen, Germany

Mathias Klinghammer Department of Experimental Psychology, Justus-Liebig-University, Giessen, Germany

Katja Fiehler Department of Experimental Psychology, Justus-Liebig-University, Giessen, Germany

In this study, we investigated the influence of gaze and prior knowledge about the reach target on the use of allocentric information for memory-guided reaching. Participants viewed a breakfast scene with five objects in the background and six objects on the table. Table objects served as potential reach targets. Participants first encoded the scene and, after a short delay, a test scene was presented with one table object missing and one, three, or five table objects horizontally shifted in the same direction. Participants performed a memory-guided reaching movement toward the position of the missing object on a blank screen. In order to examine the influence of gaze, participants either freely moved their gaze (*free-view*) or kept gaze at a fixation point (*fixation*) throughout the trial. The effect of prior knowledge was investigated by informing participants about the reach target either before (*preview*) or after (*nonpreview*) scene encoding. Our results demonstrate that humans use allocentric information for reaching even if a stable retinal reference is available. However, allocentric coding of reach targets is stronger when gaze is free and prior knowledge about the reach target is missing.

Introduction

Previous research suggests that people use two broad classes of reference frames to plan and execute reaching movements: an egocentric reference frame representing the absolute position of an object with respect to the observer and an allocentric reference frame that codes the position of an object relative to other objects in the environment (Battaglia-Mayer, Caminiti, Lacquaniti, & Zago, 2003; Colby, 1998). Regarding egocentric reference frames, targets for reaching are predominantly coded and updated with respect to gaze, i.e., in a gaze-centered or retinal coordinate system (Crawford, Henriques, & Medendorp, 2011). There is converging evidence for a combined use of egocentric and allocentric reference frames, which has been shown to result in increased accuracy and precision of reaching movements to present and remembered targets (Byrne, Cappadocia, & Crawford, 2010; de Grave, Brenner, & Smeets, 2004; Diedrichsen, Werner, Schmidt, & Trommerhäuser, 2004; Krigolson, Clark, Heath, & Binsted, 2007; Krigolson & Heath, 2004; Obhi & Goodale, 2005; Schütz, Henriques, & Fiehler, 2013; Schütz, Henriques, & Fiehler, 2015).

Previous work from our group demonstrated that targets for memory-guided reaching are coded with respect to other objects in the environment, i.e., in an allocentric reference frame (Fiehler, Wolf, Klinghammer, & Blohm, 2014, Klinghammer, Blohm, & Fiehler, 2015; Klinghammer, Blohm, & Fiehler, 2017; Klinghammer, Schütz, Blohm, & Fiehler, 2016). For example, in the study of Fiehler et al. (2014), participants were presented with a naturalistic breakfast scene that contained six objects on a table (table objects) and three objects in the environment (background objects). After scene encoding, a 2-s delay occurred followed by a test scene with one table object missing and one, three, or five of the other table objects shifted to the left or to the right. Participants were instructed to reach to the position of the missing object on a blank screen. We found that reaching movements were influenced by the object shifts as reaching endpoints were misplaced in the shift direction and increased with the number of shifted objects. Moreover, the object's contribution to allocentric coding significantly depended on its task-relevance (Klinghammer et al., 2015) and its spatial reliability (Klinghammer et al., 2017). Despite the fact that we used

Citation: Lu, Z., Klinghammer, M., & Fiehler, K. (2018). The role of gaze and prior knowledge on allocentric coding of reach targets. *Journal of Vision*, *18*(4):22, 1–13, https://doi.org/10.1167/18.4.22.

https://doi.org/10.1167/18.4.22 Received July 31, 2017; published April 25, 2018 ISSN 1534-7362 Copyright 2018 The Authors

complex, naturalistic scenes instead of impoverished stimulus displays to increase ecological validity, the applied tasks were rather unnatural as (a) gaze had to stay fixed after scene encoding until the end of the reach and (b) the reach target was unknown during scene encoding. The question arises whether and how allocentric information is used for memory-guided reaching in more everyday situations when gaze is unrestricted and knowledge about the movement target is available.

Eye movements provide important information about processing and retaining scene information. For example, previous studies from our group demonstrated that during scene encoding participants tend to fixate task-relevant objects while ignoring task-irrelevant ones (Fiehler et al., 2014; Klinghammer et al., 2015). Such task-dependent eye-movement patterns have been found in a series of perceptual and motor tasks (DeAngelus & Pelz, 2009; Land & Hayhoe, 2001; Rothkopf, Ballard, & Hayhoe, 2007; Triesch, Ballard, Hayhoe, & Sullivan, 2003), supporting top-down control of eye movements. Moreover, eye movements seem to influence working memory maintenance as the spatial memory span was found to be impaired when saccade planning and execution was prevented during information encoding or retention (Pearson, Ball, & Smith, 2014).

In the present study, we investigated how gaze and prior knowledge about the reach target influence the use of allocentric information for memory-guided reaching. To this end, we adapted the paradigm of our previous studies (Fiehler et al., 2014; Klinghammer et al., 2015; Klinghammer et al., 2017) in two ways. First, we varied *gaze* by instructing participants to either move their eyes freely without restrictions or keep gaze at a central fixation point throughout the trial, i.e., from scene encoding until the end of the reach. If a stable fixation point is available, participants could rely on a precise egocentric, retinal target representation. This should result in a decreased use of allocentric information than when gaze is unrestricted, and thus, a precise retinal reference is missing. Moreover, it has been shown that eye movements introduce noise into the egocentric spatial representation (Byrne & Crawford, 2010), which may result in a stronger weighting of allocentric representations. Second, we manipulated the *prior knowledge* about the reach target. Half of the participants were informed about the reach target before the start of each trial whereas the other half had to identify the missing object from a test scene briefly presented before the start of the reach; i.e., participants did not have information about the reach target during scene encoding, analogous to our previous studies (Fiehler et al., 2014; Klinghammer et al., 2015; Klinghammer et al., 2017). Prior knowledge has been suggested to generate an abstract visual representation, which could be retained in memory and used to guide subsequent eye movements (Castelhano & Henderson, 2007). If participants knew exactly which object to reach to, allocentric information from other objects in the scene should no longer be necessary as they could solely encode the position of the reach target relative to an egocentric reference point. This should result in a decreased use of allocentric information than when the reach target is unknown during scene encoding, and thus, a spatial representation of all potential targets has to be created. In addition, we expect differences in fixation behavior when gaze is free with more fixations at the reach target during scene encoding as well as during retention when the target is known.

Methods

Participants

We recorded data from 27 participants with normal or corrected-to-normal vision. Seven participants were excluded from data analyses: one participant because the reaching movements were not recorded due to technical difficulties and six participants because more than 25% of their trials did not meet the fixation criteria (see Data reduction and statistical analysis). Hence, the final sample consisted of 20 participants (10 male, 10 female) who were all right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971; $M = 80$, $SD = 19$) and ranged in age from 20 to 32 years ($M = 25$, $SD = 3$). Participants gave informed written consent and received course credit or financial compensation. The experimental procedures were approved by the local ethics committee of the University of Giessen in compliance with the Declaration of Helsinki (2008).

Materials

Participants were presented with 3-D–rendered images showing a breakfast scene. These images were created with SketchUp Make 2013 (Trimble Navigation Ltd, Sunnyvale, CA) and rendered with Indigo Renderer 3.6.26 (Glare Technologies Ltd) with a resolution of 3,562 × 2,671 pixels. The breakfast scenes consisted of six table objects, including an apple, a butter dish, an espresso cooker, an egg cup, a coffee mug, and a vegemite jar. The table objects were placed on a brown table located 90 cm in front of a gray wall. In addition, there were five background objects including a chair, a floor lamp, a painting on the wall, a plant, and the table that surrounded the table objects in the scene. We took these objects from the open access

Object	Height (visible)	Width	Distance to camera
Apple	6.90	5.70	variable
Butter	4.91	8.40	variable
Egg	7.45	4.92	variable
Espresso cooker	15.10	8.47	variable
Vegemite	11.44	6.72	variable
Mug	9.62	7.90	variable
Table	8.48	78.00	154.00
Plant	51.28	37.52	212.50
Painting	25.63	42.75	232.52
Chair	15.40	30.48	193.50
Lamp	54.40	24.53	212.50

Table 1. Maximum height, width, and distance to camera of objects in the scene in centimeters, based on the actual properties in SketchUp. *Notes*: Table objects had no fixed distance to the camera as they were randomly placed on one of three different depth lines. However, the reported size relates to their absolute values in SketchUp. Some background objects were not fully visible due to an overlap with other background objects or partial cutting by the image borders. In that case, the absolute size of the actually visible object part is reported here.

online 3-D gallery of SketchUp. Their properties are summarized in Table 1. Although the background objects were always placed at the same positions, the six table objects appeared at one of three horizontal depth lines that were placed starting 19.5 cm from the front table edge and distributed equally on the table. The placement of the table objects followed three criteria: (a) At each depth line, there were at least one and at most three objects; (b) the distance from the objects to the table edge was defined in a way that the object displacement would never lead to the object standing at the table edge or in the air; and (c) in case an object was occluded by other objects, the occluded part would never exceed 20%. In this way, we created nine encoding images with nine different table object position arrangements. For each of the encoding images, we created six test images for the baseline in which one table object was missing. Furthermore, based on each test image of the baseline, we created three more test images with one, three, or five table objects shifted horizontally by 3 cm at each depth line, which corresponded to a change in visual angle between 3.08° and 4.08° ($M = 3.61° \pm 0.37°$) either to the left or to the right. Variations in the horizontal displacement arose from the fact that objects were placed at different depth lines relative to the virtual camera position. Hence, similar physical shifts of objects at different depth lines in 3-D space would result in different displacements in the 2-D image.

All in all, we created 225 images with nine encoding images and 216 test images, including 54 with no shift at all (baseline), 54 images with one table object shift (*Shift 1*), 54 with three table object shifts (*Shift 3*), and 54 with five table object shifts (*Shift 5*). In addition, from each of the encoding images, we created a scrambled image that was made up of 768 randomly arranged squares and presented them for 200 ms after the offset of the encoding images in order to mask them.

Apparatus

We presented the stimuli on a 19-in. (40.5 × 30 cm) CRT monitor (Iiyama MA203DT) with a resolution of 1,280 × 960 pixels and a refresh rate of 85 Hz. A black cardboard frame (70 × 50 cm) was attached to the monitor in order to reduce the influence of a high-contrast frame around the scene. Participants sat at a desk with their head stabilized on a chin rest. The distance from the eyes to the center of the screen was about 47 cm. Reaches were recorded with an Optotrak Certus (NDI, Waterloo, Canada) tracking system with a sampling rate of 150 Hz. Therefore, an infrared marker was attached to the fingernail of the participant's right index finger. A decimal keyboard was placed in front of the participant with the start button aligned to the chin rest and the center of the screen with a distance of 24 cm from the screen. In order to control for correct fixation behavior, eye movements were recorded with an Eyelink II System (SR research, Osgoode, Canada) with a sampling rate of 500 Hz. Presentation 16.5 (Neurobehavioral Systems, Inc, Berkeley, CA) was used to program and run the experiment as well as to control the devices.

Experimental paradigm and procedure

We applied a mixed design, which consisted of two within-subject factors and one between-subjects factor. The first within-subject factor was the number of horizontally shifted table objects (*shift number*) with three levels, including shifts of one, three, or five table objects (*Shift 1*, *Shift 3*, *Shift 5*). The second within-subject factor was gaze (*gaze*) with two levels (*fixation* vs. *free-view*). In the *fixation* condition, participants were instructed to fixate a fixation point from the beginning until the end of the trial. In the *free-view* condition, they were free to move their gaze. Due to differences in the trial structure (see below), *prior knowledge* was implemented as a between-subjects factor. Participants were either informed about the reach target before the presentation of the encoding scene (*preview*), i.e., prior knowledge was available, or they received no information about the reach target before scene encoding (*nonpreview*), i.e., no prior knowledge was available. Within each group, the *fixation* and *free-view* conditions were blocked, and the two blocks were counterbalanced across participants.

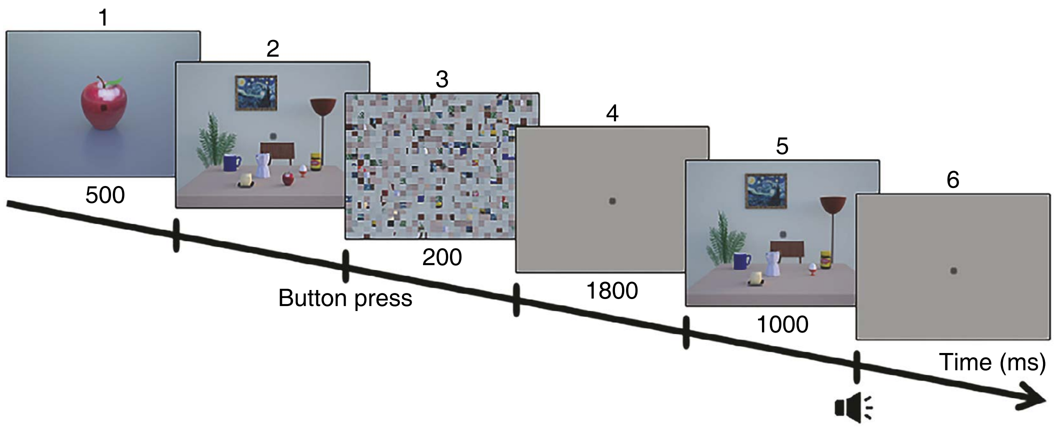

Figure 1. Trial scheme of an example trial of the *fixation* condition in the *preview* group. Participants were instructed to keep gaze at the fixation dot in the center of the screen from trial start to end. (1) First, participants viewed the target object for 500 ms. (2) Then, the encoding scene was presented without time restriction. (3) After a button press, a scrambled version of the encoding scene appeared for 200 ms, followed by (4) a delay with a gray screen which lasted for 1,800 ms. (5) Thereafter, the test scene was presented for 1,000 ms before (6) a tone prompted participants to reach to the position of the remembered target onto a gray screen.

Within each block, the *Shift 1*, *Shift 3*, and *Shift 5* conditions were presented in pseudorandomized order.

Before the start of each trial, participants had to fixate a fixation dot in the center of a blank screen and then pressed a button with their right index finger in order to initiate the drift correction for the eye-tracking device. The procedure of an example trial in the *fixation* condition for the *preview* group is illustrated in Figure 1. Participants were instructed to keep gaze at the fixation dot presented in the center of the screen from the start until the end of the trial. Each trial started with the brief presentation (500 ms) of an image of the target object. Then, the encoding scene was presented without time restriction (encoding phase). After participants pressed a button a scrambled image of the encoding scene appeared for 200 ms to avoid afterimages, followed by a retention delay with a blank background (1,800 ms, delay phase). Afterward, the test scene was presented (1,000 ms) with one of the table objects missing (= reach target) and one (*Shift 1*), three (*Shift 3*), five (*Shift 5*), or none (baseline) of the remaining table objects shifted. The presentation length of the test scene was chosen so that participants had enough time to identify the missing target object based on previous piloting. After the presentation of the test scene, a short sound (go signal) was played, prompting participants to reach to the missing table object on a blank screen.

Participants were instructed to perform the reaching movement as accurately and as natural as possible (not too fast and not too slow) and to touch the monitor at the position of the remembered reaching target. Whenever they were unsure about the target location, they had to reach to a marked location at the lower right edge of the monitor. After the reach, a black screen with a white ring was presented, prompting participants to return their finger to the starting position and start the next trial. Trials that participants had marked as invalid or in which they had started the reach before the go signal were repeated at the end of the experiment.

The procedure of the other conditions slightly differed from the *fixation* condition for the *preview* group. In the *fixation* condition for the *nonpreview* group, participants were also instructed to fixate the fixation dot throughout the trial, but they were not informed about the reach target before scene encoding. Here, the reach target had to be identified from the test scene (= missing table object). In the *free-view* condition, participants were free to move their eyes throughout the trial and either received information about the reach target before scene encoding (*preview* group) or had to identify the reach target from the test scene (*nonpreview* group). Note that the fixation dot was present in each encoding and test image in order to keep the visual input constant.

Data reduction and statistical analysis

We preprocessed data with MATLAB R2015b (MathWorks, Natick, MA) and preformed inferential statistics with R 3.2 (R Development Core Team, www.r-project.org). The alpha level was set at 0.05, and Bonferroni–Holm correction was applied to correct for multiple testing if necessary. The assumption of sphericity for the ANOVAs was tested with Mauchly's sphericity test, and Greenhouse–Geisser correction was applied if the assumption was violated.

First, we removed all the trials that were repeated (135 trials = 1.54%) from data analysis. Second, we inspected the reach movement data for each trial. Reaching onsets were defined as the time point when participants released the response button whereas offsets were defined as the first time point for which the velocity of the index finger was less than 20 mm/s and the distance from the index finger to the screen was less than 3 cm. We rejected data in trials that showed no hand movements or contained less than 20 Optotrak data samples or when reach offset criteria were not met (303 trials, 3.45%). Then, we examined the eye-tracking data in the *fixation* condition and excluded trials in which participants' fixation position deviated more than 2.5° from the fixation dot (391 trials = 4.46%). Finally, we extracted reach endpoints at the time of each reaching offset and discarded trials in which the reach endpoints deviated more than 2.5 *SD* in the vertical or horizontal direction from the group means for each object shift direction in each condition and each group (192 trials = 2.19%). In total, 7,754 valid trials from originally 8,775 trials (88.36%) remained after data reduction and were entered into statistical data analyses.

We analyzed participants' fixation behavior in the *free-view* condition during scene encoding and retention to investigate how the fixation pattern changes depending on prior knowledge about the reach target. To this end, we created target areas by drawing rectangular boxes centered on the table objects (potential reach targets). The height and width of the boxes corresponded to the maximum height and width of the respective table object (Table 1). Then we enlarged each box to 150% of its original size to cover fixations both at the table object and close to the table object. We calculated the proportion of fixations inside the target area as the ratio of the number of fixations in the target area to the total number of fixations during the encoding phase in a respective trial. Proportion of fixations inside the target area was entered into a 2×2 mixed ANOVA with *trial phase* (*encoding* vs. *delay*) as a within-subject factor and *prior knowledge* (*nonpreview* vs. *preview*) as a between-subjects factor. If participants make use of their prior knowledge of the reach target, we expect more fixations in the target area in the *preview* compared to the *nonpreview* group during scene encoding as well as during retention as reflected in a main effect of *prior knowledge*.

As the encoding phase was self-paced, we controlled for differences in the scene encoding time. Therefore, we conducted a $3 \times 2 \times 2$ mixed ANOVA with encoding time (start of scene presentation until participant's button press) as a dependent variable, *shift number* (*1*, *3*, *5*) and *gaze* (*fixation* vs. *free-view*) as within-subject factors, and *prior knowledge* (*nonpreview* vs. *preview*) as a between-subjects factor.

Second, we calculated the constant reaching error for each *shift number* by subtracting the reaching error values in the baseline from the values in the *Shift 1*, *Shift 3*, and *Shift 5* conditions. These baseline-corrected reach endpoints were then averaged and compared to the actual reach target positions. If there is an influence of allocentric information on reaching, then we expect reach endpoints to systematically deviate from the actual target position in the direction of object shifts.

Third, we determined the allocentric weight by comparing the observed baseline-corrected reaching errors with the maximal expected reaching error (MERE). The MERE was estimated by assuming that the reach endpoint errors were equal to the amount of the physical displacement of the objects when participants solely relied on the allocentric information to localize objects in space and was calculated by averaging the amount of displacement of the shifted objects for each image (Klinghammer et al., 2015). For example, if three out of the five table objects were shifted by 3 cm to the left, the MERE should be the sum of the displacements divided by the number of shifted objects, resulting in 3 cm left from the original reach target position; if all five table objects were shifted by 3 cm to the left, the MERE should also be 3 cm left from the original reach target position. The allocentric weight is then defined as the slope of a linear regression of the observed reach endpoints and the MERE, which was calculated for each participant by having the MERE as the independent variable of the linear regression and the observed baseline-corrected horizontal reaching error as the dependent variable. A slope of one would indicate that the baseline-corrected reaching error equates the MERE, i.e., participants completely rely on the allocentric information given by the shifted objects, while a slope of zero would indicate no use of allocentric information of the shifted objects (equal to baseline). First, we tested if the allocentric weights in each condition and group significantly differed from zero by using two-sided, one-sample *t* tests. If allocentric information is used for memory-guided reaching, allocentric weights should be significantly greater than zero (baseline), i.e., reach endpoints systematically deviate in the direction of object shifts. In order to assess how *gaze* and *prior knowledge* influence allocentric coding of reach targets, we conducted a $3 \times 2 \times 2$ mixed ANOVA with allocentric weight (= slope) as the dependent variable, *shift number* (*1*, *3*, *5*) and *gaze* (*fixation* vs. *free-view*) as within-subject factors, and *prior knowledge* (*nonpreview* vs. *preview*) as a between-subjects factor. As we found significant differences in the encoding time for *gaze* and *prior knowledge* (see Results), we further controlled for the influence of *encoding time* by adding it as a between-group covariate to the three-way ANOVA. In line with our previous studies (e.g., Fiehler et al., 2014), we expect allocentric weights to increase with an increasing number of object shifts. According to our hypotheses on the effects of *gaze* and *prior knowledge*, we expect higher allocentric weights in the *free-view* than the *fixation* condition as no stable retinal reference point is available. Moreover, allocentric weights should be higher in the *nonpreview* than the *preview* group as the reach target is unknown and all table objects need to be spatially encoded. This result pattern should be reflected in a main effect of shift number, a main effect of gaze, and a main effect of prior knowledge.

Finally, we conducted an exploratory analysis of participants' reaching trajectories to examine how the object shifts influenced the horizontal deviations of the reaching trajectories depending on *gaze* and *prior knowledge*. We excluded trials with more than five missing values per trajectory (524, 6.8%) or omitted the missing values in trials with five or fewer missing values per trajectory (354, 4.6%). Such omissions do not affect movement durations because they were calculated with raw time stamps from the onset and offset of movements. However, they slightly influence the curvature measure. More importantly, such influences are trifling because the frame omission only involved 4.6% of the trials and at most ~35 ms (five frames) out of an averaged movement duration of ~700 ms. In order to investigate the time course of the reach trajectories, we divided the reach durations into 11 fractions. The first and last fractions contain only 5% of the reaching time (0%~5% and 95%~100%), and each of the other nine time fractions contain 10% of the reaching time (5%~15%, 15%~25%, etc.). In the first time fraction, the finger positions were always close to the same starting point and in the last time fraction always close to the target location. Therefore, we only took the nine time fractions in the middle into consideration and averaged across the participants' horizontal coordinates of the reaching trajectories in each time fraction.

Results

Before investigating how the reaching errors and allocentric weights were influenced by the *shift number*, *gaze*, and *prior knowledge*, we examined the participants' eye fixations during the encoding and the delay phase.

Fixations

As a sanity check, we first examined whether participants' fixation behavior differed between the *fixation* and the *free-view* condition. As can be seen from the heat maps in Figure 2, in the *fixation* condition, both *nonpreview* and *preview* groups followed the instructions and kept gaze at the fixation dot, and they showed a more widespread fixation pattern distributed across the table objects in the *free-view* condition during scene encoding and retention.

In order to investigate how prior knowledge influences fixation behavior, we analyzed the *free-view* condition. We first examined the fixation distribution during the encoding and delay phases in the *preview* group in which the reach target was defined before scene encoding and in the *nonpreview* group in which the reach target was unknown during encoding. As can be seen from the sample heat maps in Figure 3, during the encoding phase, participants in the *nonpreview* group tended to equally scan all table objects whereas the *preview* group mainly fixated at the reach target. During the delay phase in which participants viewed a blank screen, the *nonpreview* group primarily directed gaze at the center of the screen whereas the *preview* group kept gaze at the position of the reach target.

This was supported by the proportion of fixations into the reach target area. Overall, 36.6% ($SD = 29.4\%$) of all fixations fell into the target area in the *preview* group, and only 10.7% ($SD = 14.3\%$) of all fixations fell into the target area in the *nonpreview* group. During the encoding phase, 29.5% ($SD = 23.8\%$) of all fixations were directed in the target area whereas during the delay phase the proportion decreased to 17.9% ($SD = 28.2\%$). Accordingly, the two-way ANOVA revealed a main effect of *prior knowledge*, $F(2, 18) = 25.409$, $p < 0.001$, $\eta^2 = 0.52$, and a main effect of *trial phase*, $F(2, 18) = 15.574$, $p \leq 0.001$, $\eta^2 = 0.18$. The interaction of *prior knowledge* and *trial phase* fell short of significance, $F(2, 18) = 0.140$, $p = 0.71$, $\eta^2 = 0.002$. Differences in the encoding time were taken into account by calculating the proportion of fixations inside the target area (see Methods). Consistent with our hypothesis, these findings indicate that participants made use of the information about the reach target by showing more fixations into the target area during scene encoding and retention when the target is known.

Encoding time

Participants were free to choose how long they wanted to view the encoding scene (see Experimental paradigm and procedure). In order to test whether *shift number*, *gaze*, or *prior knowledge* systematically influenced the encoding time, we conducted a 3 × 2 × 2 mixed ANOVA with *shift number* (*1*, *3*, *5*) and *gaze* (*fixation* vs. *free-view*) as within-subject factors and *prior knowledge* (*nonpreview* vs. *preview*) as a between-subjects factor. We found a main effect of *gaze*, $F(1, 18) = 12.83$, $p = 0.002$, $\eta^2 = 0.09$, and *prior knowledge*, $F(1,$

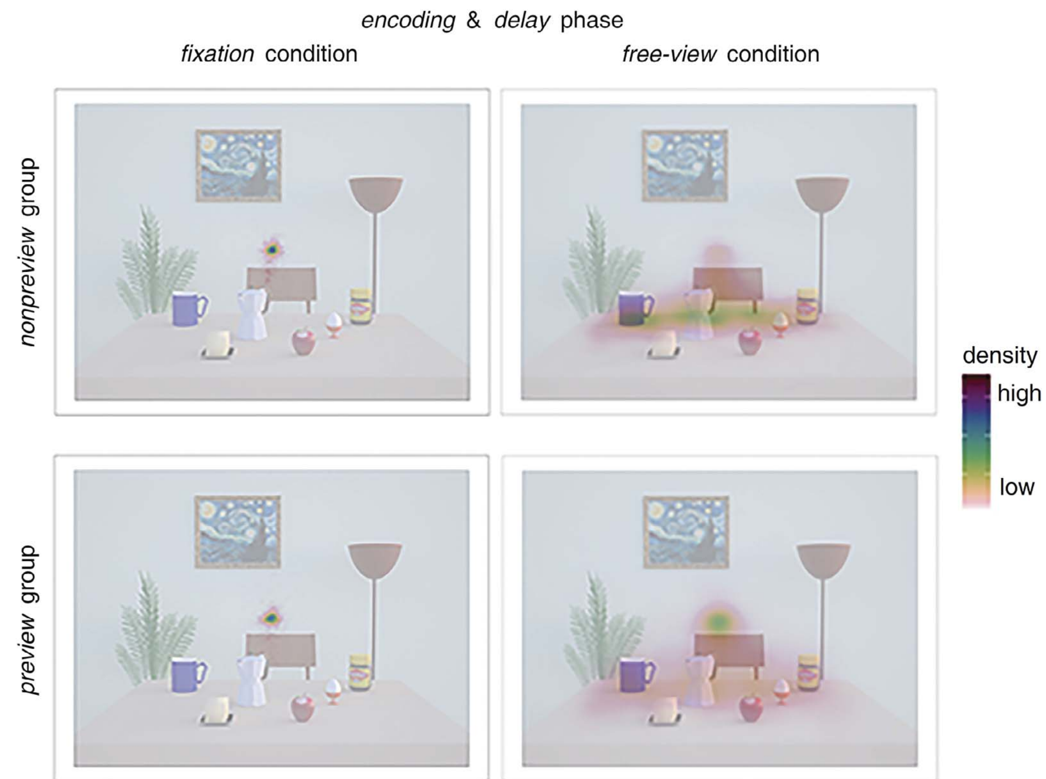

Figure 2. Heat maps of participants' fixation behavior in the *fixation* and *free-view* conditions and the *nonpreview* and *preview* groups, averaged across all nine arrangements, six targets, 20 participants, and the *encoding* and *delay* phases. For illustration purposes, we plotted the fixation density maps on the respective scene of the encoding phase.

18) = 13.16, $p = 0.002$, $\eta^2 = 0.38$, showing shorter encoding times in the *fixation* than the *free-view* condition and in the *preview* group than the *nonpreview* group.

Reaching errors and allocentric weights

Consistent with our previous findings (e.g., Fiehler et al., 2014; Klinghammer et al., 2015), reaching errors increased with the number of shifted objects, showing no effect for one table object shift but a clear effect when all remaining five table objects were shifted with three table object shifts in between (Figure 4). When examining the reaching errors averaged across *shift number* and *prior knowledge*, they clearly deviated in the direction of table object shifts in the *free-view* condition but less so in the *fixation* condition. Moreover, reaching errors averaged across *shift number* and *gaze* deviated toward the shifted table objects in both the *nonpreview* and the *preview* groups with a stronger effect for the *nonpreview* group.

In order to statistically test the influence of *shift number*, *gaze*, and *prior knowledge* on allocentric weights, we first tested the allocentric weights of each condition and each group against zero (= baseline). As shown in Table 2 and Figure 5, we found that allocentric weights significantly differed from zero in the *free-view* condition for both *nonpreview* and *preview* groups when three or five table objects were shifted and in the *fixation* condition for the *nonpreview* group when five table objects were shifted. We observed no difference from zero in the *fixation* condition for the *preview* group in which gaze was fixed and the reach target was known even when five objects were shifted.

Second, we calculated a three-way mixed ANOVA with the factors *shift number*, *gaze*, and *prior knowledge*, and it revealed no main effect of *prior knowledge*, $F(1, 18) = 0.52$, $p = 0.48$, $\eta^2 = 0.02$, but a main effect of *shift number*, $F(2, 36) = 43.54$, $p < 0.001$, $\eta^2 = 0.32$, and *gaze*, $F(1, 18) = 11.71$, $p = 0.003$, $\eta^2 = 0.08$. The main effect of *shift number* confirms our descriptive findings on reach endpoints (Figure 4) by showing an increase in allocentric weights with the number of shifted objects. The main effect of *gaze* supports our hypothesis that allocentric weights are higher in the *free-view* than the *fixation* condition. In contrast to our hypothesis, we did not find a main effect of *prior knowledge*, but an interaction between *prior knowledge* and *shift number*, $F(2, 36) = 9.41$, $p < 0.001$, $\eta^2 = 0.09$, indicating that the increase of allocentric weights with the number of shifted objects was more pronounced in the *nonpreview* than the

Figure 3. Sample heat maps of participants' fixation behavior during the *encoding* and *delay* phases in the *free-view* condition. The espresso cooker was the reach target. Please note that a blank screen (not shown here) was presented during the *delay* phase. For illustration purposes, we plotted the fixation density maps on the respective scene of the *encoding* phase.

preview group. In order to test whether the reported effects are caused by changes in encoding time, which varied with *gaze* and *prior knowledge* (see Results above), we recalculated the three-way mixed ANOVA with *encoding time* as a covariate. We found the same significant effects as reported above: main effect of *shift number*, $F(2, 36) = 41.18$, $p = < 0.001$, $\eta^2 = 0.37$; main effect of *gaze*, $F(1, 18) = 4.64$, $p = 0.045$, $\eta^2 = 0.04$; interaction between *prior knowledge* and *shift number*, $F(2, 36) = 8.44$, $p \leq 0.001$, $\eta^2 = 0.11$.

Reaching trajectories

Figure 6A illustrates the reaching trajectories averaged across *gaze* and *prior knowledge*. In line with the

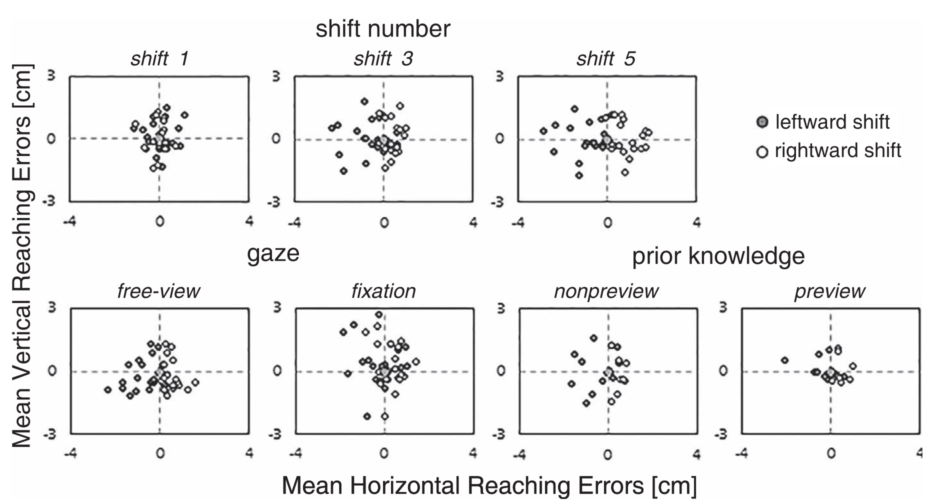

Figure 4. Mean horizontal and vertical baseline-corrected reaching errors (in centimeters) of every participant for each condition and each group. Leftward object shifts are depicted in white, rightward objects shifts in gray.

Prior knowledge	Gaze	Shift number	Range	*M*	*SD*	*t* test results
Nonpreview	*free-view*	*1*	−0.0770 to 0.0706	0.0100	0.0538	$t(9) = 0.58864$, $p = 0.571$
Nonpreview	*free-view*	*3*	−0.0034 to 0.6139	0.2423	0.1773	$t(9) = 4.3208$, $p < 0.001$*
Nonpreview	*free-view*	*5*	0.0267 to 0.9398	0.4408	0.2569	$t(9) = 5.5385$, $p < 0.001$*
Nonpreview	*fixation*	*1*	−0.3694 to 0.2592	−0.1277	0.1806	$t(9) = -2.235$, $p = 0.105$
Nonpreview	*fixation*	*3*	−0.2525 to 0.5306	0.1733	0.2302	$t(9) = 2.3798$, $p = 0.124$
Nonpreview	*fixation*	*5*	−0.0376 to 0.6665	0.2885	0.2298	$t(9) = 3.9702$, $p = 0.013$*
Preview	*free-view*	*1*	−0.0142 to 0.2645	0.0694	0.0789	$t(9) = 2.7837$, $p = 0.085$
Preview	*free-view*	*3*	0.0252 to 0.6384	0.1828	0.1691	$t(9) = 3.4181$, $p = 0.038$*
Preview	*free-view*	*5*	0.0889 to 0.8994	0.2622	0.2352	$t(9) = 3.5262$, $p = 0.039$*
Preview	*fixation*	*1*	−0.1532 to 0.1643	0.0162	0.0933	$t(9) = 0.5482$, $p = 0.597$
Preview	*fixation*	*3*	−0.1281 to 0.4518	0.0884	0.1559	$t(9) = 1.793$, $p = 0.213$
Preview	*fixation*	*5*	−0.1287 to 0.6825	0.1378	0.228	$t(9) = 1.9114$, $p = 0.265$

Table 2. Summary of allocentric weights for each condition and each group. *Note*: *$p < 0.05$, Bonferroni-Holm corrected.

results on reaching endpoints, reaching trajectories deviated in the direction of object shifts, and this scaled with the number of shifted objects. The more objects were shifted the stronger reaching trajectories deviated into the direction of the object shift. Figure 6B and C illustrates the reaching trajectories averaged across *shift number* for the *gaze* conditions and *prior knowledge* groups. As visible in Figure 6C, reaching trajectories showed a stronger deviation in the direction of the object shifts in the *preview* than the *nonpreview* group. In addition, reaching trajectories in the *preview* group started to deviate earlier in the *fixation* than the *free-view* condition. These descriptive results indicate that object shifts had a greater influence on participants' reaching trajectories in the *preview* than the *nonpreview* group, and this influence was earlier in the *fixation* than the *free-view* condition.

Discussion

We replicated our previous findings (Fiehler et al., 2014; Klinghammer et al., 2015; Klinghammer et al., 2017; Klinghammer et al., 2016) showing that reaching trajectories and endpoints are systematically influenced by object shifts in the environment and that this influence increases with the number of shifted objects. The allocentric weights ranged from –0.13 to 0.44, indicating that reaching endpoints were affected by up to 44% by the object shifts. The remaining percentage could be attributed to the influence of egocentric or other allocentric reference frames, e.g., the table in the scene or the frame of the monitor. The strength of the allocentric weights we observed here was comparable to the ones we obtained previously (Fiehler et al., 2014: –0.01–0.43; Klinghammer et al., 2015: –0.01–0.47). In contrast to our previous studies, participants in the

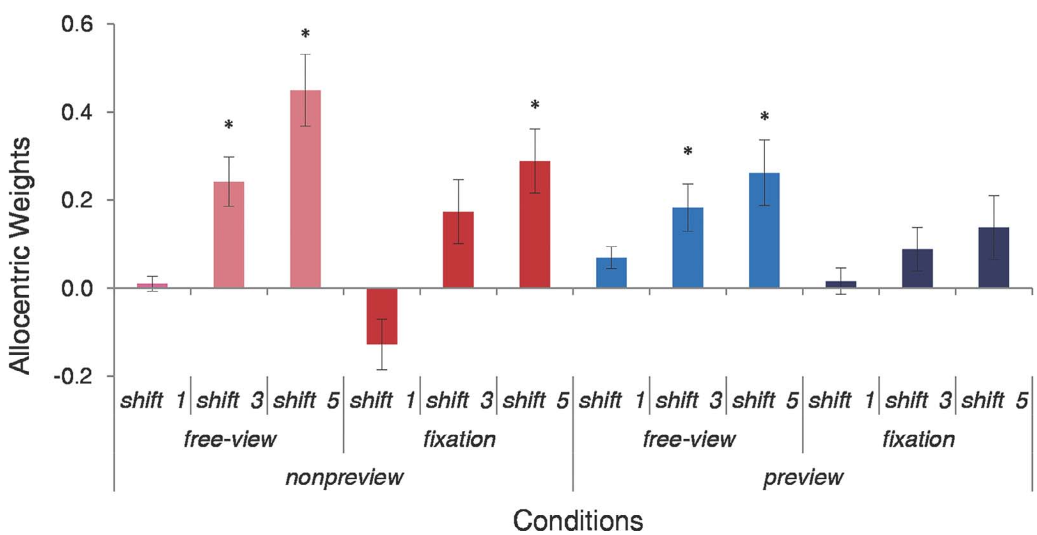

Figure 5. Allocentric weights for each condition and each group averaged across participants. Error bars represent 1 *SEM*. Asterisks indicate that allocentric weights significantly differed from zero (= baseline).

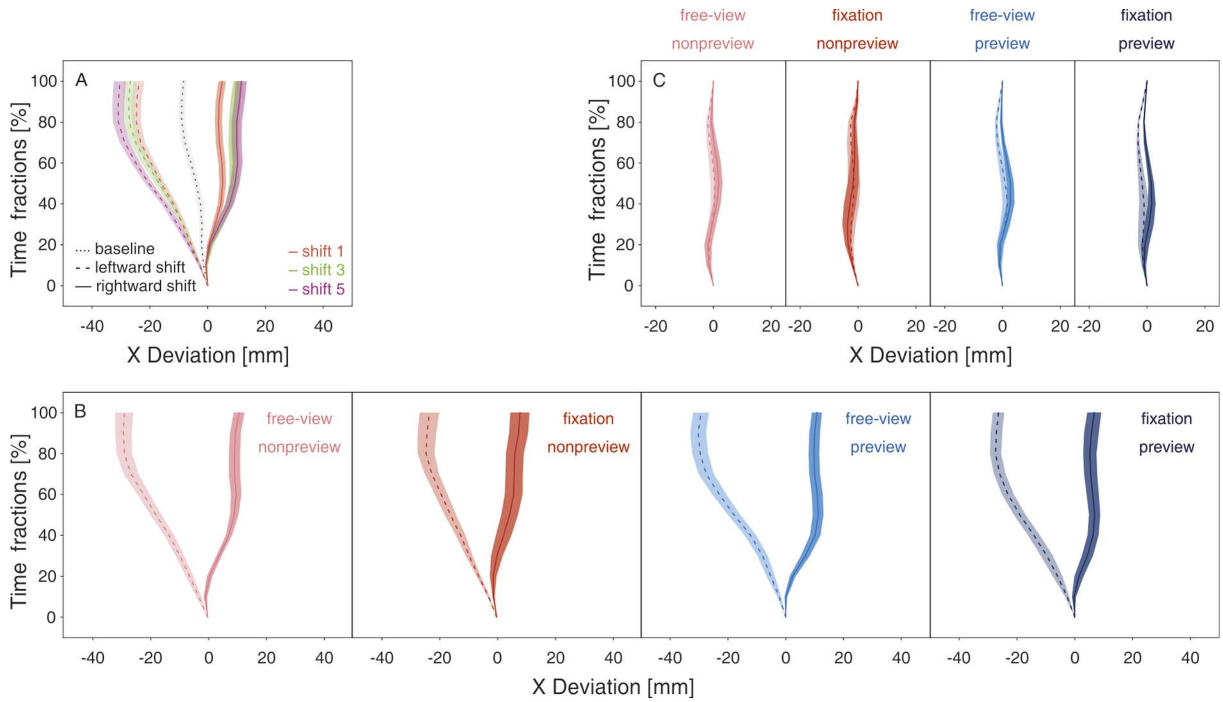

Figure 6. Mean trajectories of all participants plotted as the horizontal deviation on the *x*-axis (parallel to the screen) against the proportions of reaching duration for leftward (dashed line) and rightward (solid line) object shifts. All trajectories are scaled to the same starting point. (A) Trajectories for the baseline and *shift number* averaged across *gaze* and *prior knowledge*. (B) Trajectories for *gaze* and *prior knowledge* averaged across *shift number*. (C) Trajectories for *gaze* and *prior knowledge* averaged across *shift number,* rotated to the same endpoint.

free-view condition were not only allowed to freely move their eyes during the encoding phase, but also during the delay, test, and reaching phases. However, we obtained similar allocentric weights, indicating that gaze behavior during scene encoding primarily determines the use of allocentric information for reaching.

Gaze behavior

We found a smaller influence of allocentric information on reaching endpoints when gaze was fixed than when gaze was free. Following Filimon's (2015) argumentation, the fixation dot may have provided a stable retinal reference point during scene encoding, delay, and reaching, leading to a precise retinal target representation and, thus, eliminating the influence of allocentric information on reaching. Our results do not support this claim. We found an increase in allocentric weights with the number of shifted objects irrespective of whether gaze was free or fixed. This suggests that participants still use an allocentric reference frame for memory-guided reaching even if they can solely rely on a stable egocentric (gaze-centered) reference frame. However, allocentric weights were reduced when gaze was fixed than free, indicating a weaker influence of allocentric information when a retinal reference point is available.

In line with previous studies on gaze behavior in real-world situations (DeAngelus & Pelz, 2009; Hayhoe & Rothkopf, 2011; Land & Hayhoe, 2001; Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011; Oliva, Torralba, Castelhano, & Henderson, 2003; Rothkopf et al., 2007; Triesch et al., 2003), we found that participants mainly fixated on task-relevant table objects during scene encoding and retention when gaze was free. This information is likely to be encoded into spatial working memory and retained over several seconds (Hollingworth & Henderson, 2002; Maxcey-Richard & Hollingworth, 2013). If gaze is restricted, e.g., during fixation, objects are seen in the visual periphery, leading to less precise and, thus, less reliable information than if they are foveated. Therefore, participants may have encoded comparatively limited and unprecise visual information of object locations and, thus, have retained less detailed allocentric information in spatial working memory. Accordingly, when gaze is free, the larger influence of allocentric information on reaching endpoints may be caused by increased precision of allocentric information in working memory used for reaching. In addition, free gaze may have increased the uncertainty in the egocentric spatial representation as each gaze shift introduces eye movement–induced noise (cf. Byrne &

Crawford, 2010), resulting in a stronger reliance on allocentric information.

Prior knowledge

Allocentric information influenced reaching endpoints irrespective of whether prior knowledge about the reach target was available during scene encoding. However, the influence of allocentric information was stronger when the reach target was unknown. Previous studies demonstrated that prior knowledge from a brief glimpse could generate an abstract visual representation that can be retained in working memory and used to guide subsequent eye movements (Castelhano & Henderson, 2007; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). Moreover, eye movements have been generally considered to be controlled by top-down processes that restrict fixations to task-relevant locations (DeAngelus & Pelz, 2009; Hayhoe & Rothkopf, 2011; Land & Hayhoe, 2001; Mills et al., 2011; Oliva et al., 2003; Rothkopf et al., 2007; Triesch et al., 2003). This is supported by the present findings demonstrating that participants mainly fixated either all the table objects serving as potential reach targets or the one table object that was introduced beforehand. Taking into consideration that only the visual information of previously attended task-relevant objects are retained in working memory (Hollingworth & Henderson, 2002), participants may have had a less detailed and less precise memory representation of the other objects when the reach target was known, leading to a reduced influence of allocentric coding. However, prior knowledge did not lead to a complete lack of allocentric coding as we observed allocentric weights different from baseline (no object shift) when the reach target was known and gaze was free, i.e., in situations closest to everyday behavior. This suggests that allocentric information is still used for reaching even when the task could be solely performed in an egocentric reference frame. Such a combination of egocentric and allocentric reference frames likely provides a more precise estimate of the visual target location in space.

Prior knowledge did also influence the effect of object shifts on reaching trajectories. In contrast to our results on reaching endpoints, object shifts had a greater influence on reaching trajectories when the reach target was known than when it was unknown. Previous studies have shown that people use surrounding information to adjust their reaching movements (Brenner & Smeets, 1997; Saijo, Murakami, Nishida, & Gomi, 2005; Whitney, Westwood, & Goodale, 2003). When the reach target was unknown, participants had to use the spatial information of the table objects in the test scene to correctly locate the reach target. In this way, the object shifts were taken into account during the programming of the reach and, thus, hardly affected the reaching trajectories. On the other hand, when the reach target was known, participants may have relied on the precise spatial target representation built up during scene encoding. Hence, the shifts of the surrounding table objects in the test scene may have required adjustments during the movement to correctly reach to the remembered target location. These adjustments seem to occur earlier in time the more precise the spatial target representation is, i.e., in situations when gaze is fixed rather than free.

Eye movement behavior during scene encoding and retention

Our results showed that participants applied different encoding and retention strategies depending on their prior knowledge. When the reach target was unknown, they scanned the table objects during scene encoding and then kept gaze at the screen center on the blank screen during the delay. When the reach target was known, the eyes were mainly directed to the target location and kept there during the delay. Similar findings were obtained in a previous study that demonstrated that participants spontaneously shift their gaze to the to-be-remembered locations during a blank retention interval, which led to a better change-detection performance compared to situations in which gaze was restricted (Williams, Pouget, Boucher, & Woodman, 2013). However, there are contradictory findings on whether eye movements indeed facilitate the maintenance of information in spatial working memory. By showing a reduced working memory span in conditions in which eye movements were prevented during stimulus encoding or retention, Pearson et al. (2014) suggested that oculomotor preparation serves as a rehearsal mechanism that is necessary to optimally retain a sequence of locations in working memory. On the contrary, Godijn and Theeuwes (2012) argued that overt eye movements do not benefit from the rehearsal of visuospatial information as they did not find worse memory performance when gaze was restricted. Future studies should clarify how different encoding and retention strategies affect the use of allocentric information in memory-guided reaching.

Conclusion

In this study, we found that humans make use of allocentric information in memory-guided reaching even if a stable retinal reference point is given or the reach target is known, and thus, the task could be performed solely in an egocentric reference frame. However, the influence of allocentric information on reaching depends on gaze and prior knowledge of the reach target with stronger allocentric coding when gaze is free and the reach target is undefined during scene encoding.

Keywords: reference frames, allocentric, egocentric, gaze, prior knowledge, memory-guided reaching

Acknowledgments

This research was funded by the IRTG 1901 "The Brain in Action" by the German Research Foundation (DFG). Moreover, we would like to thank Tobias Moehler, Aaron Drehmann, Lena Klever, and Immo Schuetz for their support in conducting and analyzing this study.

Commercial relationships: none.
Corresponding author: Katja Fiehler.
Email: katja.fiehler@psychol.uni-giessen.de.
Address: Department of Experimental Psychology, Justus-Liebig-University, Giessen, Germany.

References

Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., & Zago, M. (2003). Multiple levels of representation of reaching in the parieto-frontal network. *Cerebral Cortex*, *13*(10), 1009–1022.

Brenner, E., & Smeets, J. B. (1997). Fast responses of the human hand to changes in target position. *Journal of Motor Behavior*, *29*(4), 297–310.

Byrne, P. A., Cappadocia, D. C., & Crawford, J. D. (2010). Interactions between gaze-centered and allocentric representations of reach target location in the presence of spatial updating. *Vision Research*, *50*(24), 2661–2670.

Byrne, P. A., & Crawford, J. D. (2010). Cue reliability and a landmark stability heuristic determine relative weighting between egocentric and allocentric visual information in memory-guided reach. *Journal of Neurophysiology*, *103*(6), 3054–3069.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, *20*(1), 15–24.

Crawford, J. D., Henriques, D. Y., & Medendorp, W. P. (2011). Three-dimensional transformations for goal-directed action. *Annual Review of Neuroscience*, *34*, 309–331.

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6–7), 790–811.

de Grave, D. D., Brenner, E., & Smeets, J. B. (2004). Illusions as a tool to study the coding of pointing movements. *Experimental Brain Research*, *155*(1), 56–62.

Diedrichsen, J., Werner, S., Schmidt, T., & Trommershäuser, J. (2004). Immediate spatial distortions of pointing movements induced by visual landmarks. *Perception & Psychophysics*, *66*(1), 89–103.

Fiehler, K., Wolf, C., Klinghammer, M., & Blohm, G. (2014). Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Frontiers in Human Neuroscience*, *8*: 636.

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. *Frontiers in Human Neuroscience*, *9*: 648.

Godijn, R., & Theeuwes, J. (2012). Overt is no better than covert when rehearsing visuo-spatial information in working memory. *Memory & Cognition*, *40*(1), 52–61.

Hayhoe, M. M., & Rothkopf, C. A. (2011). Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), 158–166.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1):6, 49–63, https://doi.org/10.1167/3.1.6. [PubMed] [Article]

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136.

Klinghammer, M., Blohm, G., & Fiehler, K. (2015). Contextual factors determine the use of allocentric information for reaching in a naturalistic scene. *Journal of Vision*, *15*(13):24, 1–13, https://doi.org/10.1167/15.13.24. [PubMed] [Article]

Klinghammer, M., Blohm, G., & Fiehler, K. (2017). Scene configuration and object reliability affect the use of allocentric information for memory-guided reaching. *Frontiers in Neuroscience*, *11*: 204.

Klinghammer, M., Schütz, I., Blohm, G., & Fiehler, K.

(2016). Allocentric information is used for memory-guided reaching in depth: A virtual reality study. *Vision Research*, *129*, 13–24.

Krigolson, O., Clark, N., Heath, M., & Binsted, G. (2007). The proximity of visual landmarks impacts reaching performance. *Spatial Vision*, *20*(4), 317–336.

Krigolson, O., & Heath, M. (2004). Background visual cues and memory-guided reaching. *Human Movement Science*, *23*(6), 861–877.

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

Maxcey-Richard, A. M., & Hollingworth, A. (2013). The strategic retention of task-relevant objects in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 760–782.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*(8):17, 1–15, https://doi.org/10.1167/11.8.17. [PubMed] [Article]

Obhi, S. S., & Goodale, M. A. (2005). The effects of landmarks on the performance of delayed and real-time pointing movements. *Experimental Brain Research*, *167*(3), 335–344.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In Image processing, 2003. icip 2003. proceedings. 2003 international conference on (Vol. 1, pp. I-253). IEEE.

Pearson, D. G., Ball, K., & Smith, D. T. (2014). Oculomotor preparation as a rehearsal mechanism in spatial working memory. *Cognition*, *132*(3), 416–428.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14):16, 1–20, https://doi.org/10.1167/7.14.16. [PubMed] [Article]

Saijo, N., Murakami, I., Nishida, S. Y., & Gomi, H. (2005). Large-field visual motion directly induces an involuntary rapid manual following response. *Journal of Neuroscience*, *25*(20), 4941–4951.

Schütz, I., Henriques, D. Y. P., & Fiehler, K. (2013). Gaze-centered spatial updating in delayed reaching even in the presence of landmarks. *Vision Research*, *87*, 46–52.

Schütz, I., Henriques, D. Y., & Fiehler, K. (2015). No effect of delay on the spatial representation of serial reach targets. *Experimental Brain Research*, *233*(4), 1225–1235.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1):9, 86–94, https://doi.org/10.1167/3.1.9. [PubMed] [Article]

Whitney, D., Westwood, D. A., & Goodale, M. A. (2003, June 19). The influence of visual motion on fast reaching movements to a stationary object. *Nature*, *423*(6942), 869–873.

Williams, M., Pouget, P., Boucher, L., & Woodman, G. F. (2013). Visual–spatial attention aids the maintenance of object representations in visual working memory. *Memory & Cognition*, *41*(5), 698–715.