

Spatial coding for memory-guided reaching in visual and pictorial spaces

Harun Karimpur

Experimental Psychology, Justus Liebig University,
Giessen, Germany
Center for Mind, Brain, and Behavior (CMBB), University
of Marburg and Justus Liebig University, Giessen,
Germany



Siavash Eftekharifar

Centre for Neuroscience Studies, Queen's University,
Kingston, ON, Canada



Nikolaus F. Troje

Centre for Neuroscience Studies, Queen's University,
Kingston, ON, Canada
Centre for Vision Research and Department of Biology,
York University, Toronto, ON, Canada



Katja Fiehler

Experimental Psychology, Justus Liebig University,
Giessen, Germany
Center for Mind, Brain, and Behavior (CMBB), University
of Marburg and Justus Liebig University, Giessen,
Germany



An essential difference between pictorial space displayed as paintings, photographs, or computer screens, and the visual space experienced in the real world is that the observer has a defined location, and thus valid information about distance and direction of objects, in the latter but not in the former. Thus egocentric information should be more reliable in visual space, whereas allocentric information should be more reliable in pictorial space. The majority of studies relied on pictorial representations (images on a computer screen), leaving it unclear whether the same coding mechanisms apply in visual space. Using a memory-guided reaching task in virtual reality, we investigated allocentric coding in both visual space (on a table in virtual reality) and pictorial space (on a monitor that is on the table in virtual reality). Our results suggest that the brain uses allocentric information to represent objects in both pictorial and visual space. Contrary to our hypothesis, the influence of allocentric cues was stronger in visual space than in pictorial space, also after controlling for retinal stimulus size, confounding allocentric cues, and differences in presentation depth. We discuss possible reasons for stronger allocentric coding in visual than in pictorial space.

Introduction

Imagine standing in front of a painting. After watching it for a few moments, we start to understand and interpret its geometry and the spatial relationships between the depicted objects. However, our own location is ill-defined in pictorial space. We may adopt the viewpoint from which the picture was taken as a surrogate for true location, but that location is not really ours as we have no control over it. It does not change as we move. In contrast, in the visual space of the real world, our body occupies a defined location. A distinction between the two spaces, the pictorial space that is depicted on the painting and the visual space we are a part of, seems to be crucial (Goldstein, 1987; Koenderink & van Doorn, 2003; Koenderink & van Doorn, 2008; Koenderink & van Doorn, 2012; Vishwanath, Girshick, & Banks, 2005).

Pictorial representations have dominated vision research for most of its existence. However, recent behavioral results in adults (Gomez & Snow, 2017; Snow, Skiba, Coleman, & Berryhill, 2014) and children (Gerhard, Culham, & Schwarzer, 2016), as well as electroencephalogram (Marini, Breeding, & Snow, 2019) and functional magnetic resonance imaging findings

Citation: Karimpur, H., Eftekharifar, S., Troje, N. F., & Fiehler, K. (2020). Spatial coding for memory-guided reaching in visual and pictorial spaces. *Journal of Vision*, 20(4):1, 1–12, <https://doi.org/10.1167/jov.20.4.1>.



(Freud et al., 2018; Snow et al., 2011) provide evidence that the human brain processes real objects and pictures of the same objects differently. The use of pictorial representations relies on the implicit assumption that if the retinal images received from a certain viewpoint in the real world, and the retinal images received when looking at a picture are similar, the responses of the visual system are expected to be similar too. That assumption has rarely been verified. It may hold in many cases, but there are others in which it may not. The latter is expected particularly in studies that investigate spatial coding, as they often employ paradigms in which the relative direction between object and observer or the orientation of an object relative to the observer matter (Troje, 2019).

Spatial coding for action requires to map action targets onto common coordinate systems or reference frames. An overwhelming body of evidence supports early views (Arbib, 1991; Colby, 1998; Klatzky, 1998) that these fall into two categories: in an egocentric reference frame, we map objects relative to ourselves. This requires that the brain constantly updates spatial relations between us and the objects as a consequence of our movements. In an allocentric reference frame, we map objects relative to other objects (landmarks). If landmarks are stable, they allow us to reliably compute spatial relations. It is well established that we represent targets for actions in a gaze-centered (i.e., egocentric) reference frame (Crawford, Henriques, & Medendorp, 2011; Medendorp, 2011). However, allocentric information also crucially contributes to spatial coding for action. For example, although the presence of landmarks results in more accurate and less variable reaches in both online and delayed movement tasks (Krigolson & Heath, 2004; Obhi & Goodale, 2005), cue-irrelevant background information can lead to spatial distortions of pointing or reaching movements (Diedrichsen, Werner, Schmidt, & Trommershäuser, 2004; Taghizadeh & Gail, 2014). Further studies could show that the main factors driving the influence of allocentric information are landmark stability, as well as its reliability as a cue (Byrne & Crawford, 2010; Camors, Jouffrais, Cottureau, & Durand, 2015). Nevertheless, spatial coding for action requires information about the true location of the observer, emphasizing the role of egocentric reference frames in visual space.

As pointed out earlier, the true location of the observer in pictorial space is ambiguous (Troje, 2019). Although we can see the relative location of objects in the picture, we do not have a well-defined own location relative to these objects, emphasizing the role of allocentric reference frames in pictorial space. To probe for allocentric coding, Fiehler and colleagues (Fiehler, Wolf, Klinghammer, & Blohm, 2014; Klinghammer, Blohm, & Fiehler, 2015; Klinghammer, Blohm, & Fiehler, 2017; Lu, Klinghammer, & Fiehler, 2018) conducted a series of experiments in which participants

were seated in front of a monitor and viewed pictures of a breakfast scene. After visual exploration of the scene (self-paced), the objects disappeared briefly before they reappeared again with one object missing (test scene). Then the screen blanked, and participants were asked to reach to the remembered position of the missing object on the empty screen. Crucially, in the test scene, the objects on the table were subtly shifted. The main finding here was that participants' reaching endpoints systematically deviated in the direction of the object shift, indicating the use of allocentric information for reaching (cf., Byrne & Crawford, 2010). Reaching endpoints were only influenced by shifts of objects that were potential reach targets, as instructed before the experiment, and thus task-relevant. Shifts of objects that never served as reach targets did not affect reaching, even if they were near the target or the shift caused substantial changes in the scene (Klinghammer et al., 2015). Hence allocentric coding is facilitated by task-relevant allocentric cues, whereas task-irrelevant cues are mostly ignored. Further facilitation has been shown when task-relevant objects are coherently shifted in the same direction or allowed for spatial clustering (Klinghammer et al., 2017), and when gaze was free compared with fixed in the center of the scene (Lu et al., 2018). It is important to note that stimuli in all of these studies were presented on a computer screen, that is, the objects were presented in pictorial space.

In only two studies, these findings were extended into visual space by means of virtual reality (Karimpur, Morgenstern, & Fiehler, 2019; Klinghammer, Schütz, Blohm, & Fiehler, 2016). The results imply that allocentric coding is comparable in visual and pictorial space. However, a direct comparison of these studies is difficult due to several methodological shortcomings. First, they did not directly compare pictorial and visual spaces, leaving open whether and how allocentric coding differs between the two spaces. Second, they were inconsistent with respect to the stimulus set, the retinal stimulus size, gaze control, and the coherence between real-world and computer graphics. The goal of the current study was to address these points to get a deeper understanding of how allocentric reference frames are used to represent the location of reaching targets, and how that depends on whether they belong to the visual space of the observer or were presented in pictorial space.

To this end, we conducted a memory-guided reaching task in a series of three experiments. In Experiment 1, we presented a breakfast scene similar to the ones used in the studies discussed earlier in both visual space and pictorial space. Both presentation modes were implemented in virtual reality, which allowed us to control for potentially confounding differences in display conditions, such as spatial resolution, illumination, and material properties. In the Visual Space¹ condition, participants performed reaching

movements toward the remembered position of an object placed on a table in virtual reality. In the Pictorial Space conditions, the breakfast scene was presented on a monitor placed on a table in virtual reality. It is noteworthy that the breakfast scene in the Pictorial Space conditions was a projection of the breakfast scene of the Visual Space condition. To examine the influence for retinal stimulus size, we had two Pictorial Space conditions: one with a small monitor with dimensions similar to a typical 30-in. desktop widescreen monitor (Pictorial Small), and another with a large monitor (Pictorial Large) in which objects had the same retinal size as in the Visual Space condition.

Experiment 2 served as a control for other potential allocentric cues, for example, the left and right table edges, that participants could have used to represent the task-relevant objects. Finally, in Experiment 3, we aimed to control for variability in depth. This relates to a potential concern regarding differences between presentation modes. In Pictorial Space, participants reached to the surface of the monitor. In Visual Space, the objects were presented at different depth levels on a surface. This inevitably results in different reaching movements and differences in relative disparities.

All objects that we presented on the breakfast table were task-relevant (cf., Klinghammer et al., 2016), that is, we instructed participants that only these objects served as potential reach targets. We therefore expect to replicate the previous findings in virtual reality that humans encode targets for action relative to each other, that is, in an allocentric reference frame. Because of the lack of a true location of the observer in pictorial space, we further expect higher allocentric weights in the Pictorial Space conditions compared with the Visual Space condition.

Methods

Participants

For Experiment 1, we recruited 19 students of the Justus Liebig University Giessen via university email. Participants provided informed consent and received either financial compensation or course credit. We used a graded circle test of the Stereo Fly Test battery (Stereo Optical Co., Inc., Chicago, IL), and excluded one participant due to insufficient stereopsis. Therefore the final sample consisted of 18 students (13 women; M_{age} : 24.22 years) who reported normal or corrected to normal vision, and were all right-handed ($M = 86.97$, $SD = 15.50$) as assessed via the Edinburgh Handedness Inventory (EHI; Oldfield, 1971). The experimental procedures were in accordance with the principles of the Declaration of Helsinki (World Medical Association,

2013) and approved by the local ethics committee of the Justus Liebig University Giessen.

Apparatus

The setup is depicted in Figure 1A. Participants were seated in front of a table ($120 \times 140 \times 76$, length \times width \times height; cm) with a chin rest placed at the center of the table set to a height of 30 cm. The table position, as well as the position of the right index finger, were tracked by means of an 8-camera motion capture system (VICON Vero, Oxford, UK) at 100 Hz. The experiment was presented and controlled in Unity3D (Unity Technologies, San Francisco, CA, USA) and ran on a Dell Alienware computer with an Intel® Core™ i9 7980XE processor, 32GB RAM, and two NVIDIA® GeForce® GTX™ 1080Ti graphics processing units (NVIDIA Corporation, Santa Clara, CA, USA). We presented the virtual environment stereoscopically with an HTC Vive HMD (HTC Corporation, New Taipei City, Taiwan) at a resolution of 1080×1200 pixels per eye, and a refresh rate of 90 Hz. For reaching movements in the pictorial space condition we placed a monitor on the table 60 cm in front of the participants and extended the reachable surface area of the display by attaching a 0.3 cm Plexiglas® layer (Röhm GmbH, Darmstadt, Germany) that was 100 cm in width and 62 cm in height.

We attached reference markers to the top right of the table in the laboratory and used the top right of a same-sized table in virtual reality to cross-calibrate the two spaces (reality and virtual reality). This allowed us to position, for example, the surface of the monitor in the Pictorial Space conditions at the same position relative to the observer as in reality.

Stimuli

We defined three *presentation modes*, which are depicted in Figure 1B. We presented the breakfast scene in the Visual Space on the table in virtual reality and presented it in the Pictorial Space conditions on a virtual monitor in virtual reality. In the Pictorial Small condition, the monitor had a width of ~ 65 cm and a height of ~ 42 cm ($\sim 30^\circ$ in diag.). In the Pictorial Large condition, the monitor had a width of ~ 100 cm and a height of ~ 65 cm ($\sim 47^\circ$ in diag.). The dimensions of the action-relevant objects we presented in the experiment can be found in Table 1.

For every subject and target object we created random arrangements for the six objects so that their positions fall on a 50-cm wide and 40-cm deep table area while keeping a distance of at least 15 cm between objects. In the shift conditions, we chose a displacement of 4 cm (to the left and to the right) and added a

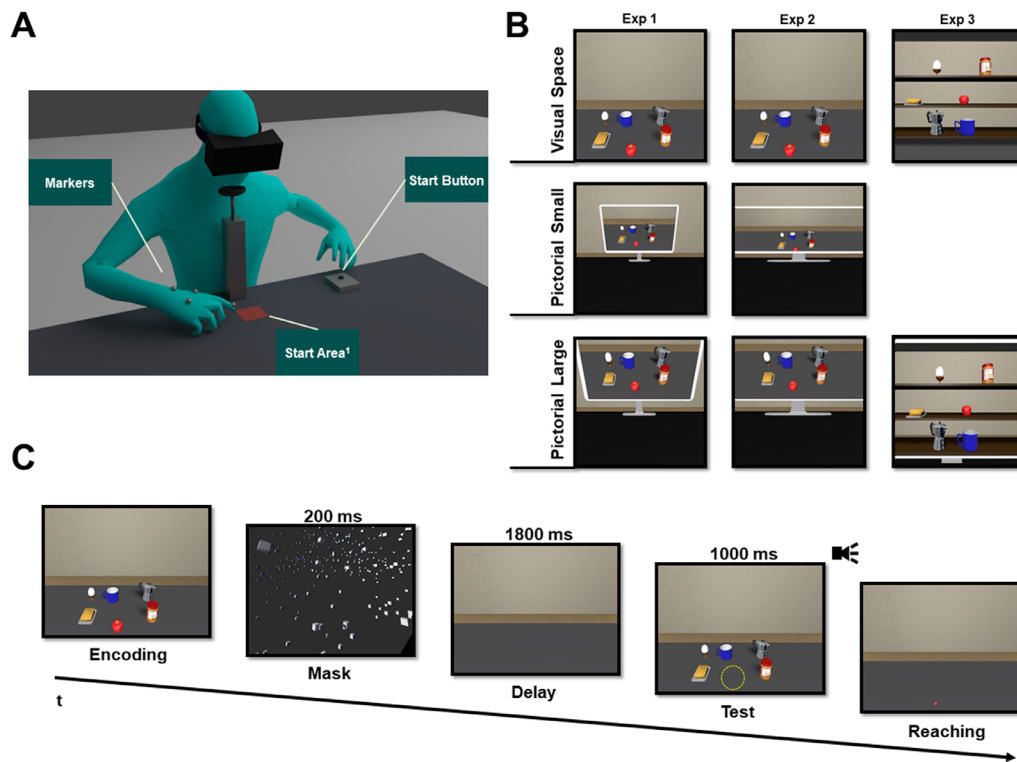


Figure 1. Setup and procedure of the study. (A) Participants position during an experiment with markers for motion tracking placed on the right hand and a start button to control the experiment. (B) Exemplary encoding scenes to demonstrate the presentation modes Visual Space, Pictorial Small, and Pictorial Large. (C) Procedure of a typical trial. The missing object (target) is marked by a yellow dotted circle in the test scene (only for illustration purposes). Please note that in Experiment 3, the procedure was identical, but the scene looked as depicted in the third column of (B).

Object	Length (cm)	Width (cm)	Height (cm)
Apple	6.5	6.4	5.2
Butter	8.5	12.6	3.7
Egg in egg cup	5.3	5.8	10.8
Espresso cooker	8.5	9.3	13.2
Mug	8.8	8.8	8.9
Peanut butter jar	7.0	7.0	12.0

Table 1. List of objects.

uniformly distributed random noise between -0.5 and 0.5 cm. Each of the six targets/arrangement was presented three times (baseline [no shift], shift left, shift right) in all three display conditions resulting in $6 \times 3 \times 3 = 54$ trials. We measured each combination three times so that each participant completed 162 trials.

We created a mask scene with 800 gray cubes that were randomly placed in a dark scene for 200 ms. In case of Visual Space, this mask was presented in the HMD. In case of Pictorial Space conditions, this mask was presented in the virtual monitor. Before the experiment started, we ensured that the participants understood

the task by practicing 10 trials with the same objects but different randomly generated arrangements. The position of the index finger was represented by a small red sphere in the virtual environment.

Procedure

The sequence of events during a single trial is depicted in Figure 1C. We presented the breakfast scene with six objects on the table. Participants were allowed to freely explore the scene without gaze restriction. To proceed, the right index finger had to be placed on the start area that was visually marked in the virtual environment. After pressing the start button (Figure 1A), a brief mask (200 ms) appeared and was followed by a delay (1800 ms) in which an empty table was presented. All objects except for the target object reappeared, remained visible for 1000 ms, and then disappeared again. After an auditory go signal, participants had to reach to the remembered target position on an empty table. The reaching endpoint was recorded when they touched the table surface with their index finger. If participants left the start area before the

go signal, they were shown a warning message and the trial was dismissed.

Data reduction and statistical analyses

We collected data from 18 participants in Experiment 1 resulting in a total of 2916 trials. For each participant, we first removed trials in which participants started their movements before the acoustic start signal ($\sim 3\%$), or when there were errors in the kinematic data, for example, owing to occluded motion capture markers ($< 1\%$). After aggregating data over the three trial repetitions, we calculated the reaching errors by subtracting reaching endpoints in the baseline condition from the reaching endpoints in the shift conditions. We analyzed the mean reaching errors separately for the three display conditions and removed trials in which the reaching endpoints deviated more than 2 *SD* from the group means ($< 1\%$).

To quantify the influence of allocentric information, we first calculated the lateral reaching errors by subtracting the reaching endpoints in the baseline (no-shift) conditions from the respective reaching endpoints in the shift conditions. If participants encoded the target object relative to the other objects on the table (allocentrically, task-relevant), a shift of the objects would cause a systematic shift of reaching endpoints in the direction of object displacement. Similarly, no systematic shift of reaching endpoints would indicate that the target was encoded relative to the observer (egocentrically), or relative to other objects or landmarks in the scene (allocentrically, task-irrelevant). Second, we calculated allocentric weights by taking the ratio of the lateral reaching errors to the average lateral displacement of the objects on the table. For instance, a reaching error of 2 cm following an average shift of 4 cm results in an allocentric weight of 0.5. A ratio of approximately 1 indicates that participants spatially encoded a given target object fully relative to the location of the other objects on the table. For the sake of completeness, it should be noted that, hypothetically, it is possible that participants overshoot with allocentric weights greater than 1, and also that reaching errors deviate in the opposite direction of object shift resulting in a negative allocentric weight.

We did not find violations of normality for the allocentric weights of each condition by visual inspection of QQ-plots and a Shapiro–Wilk test, and thus applied parametric tests. One-sample *t*-tests were conducted to test whether allocentric weights differed significantly from zero. A Mauchly test was conducted to confirm sphericity. To show whether the influence of allocentric information differed between the two presentation modes, we conducted a repeated measures analysis of variance

with *presentation mode* as within-subject factor (levels: Visual Space, Pictorial Small, Pictorial Large), and the allocentric weights as a dependent variable. Following a significant main effect, we conducted pairwise comparisons by using paired-sample *t*-tests and applied Bonferroni–Holm correction for inflated family-wise error rate. For the comparisons, we report Cohen’s d_z as a measure of effect size, which is the mean difference divided by the standard deviation of the difference (Lakens, 2013).

Changes in Experiment 2

In Experiment 2, we aimed to reduce the number of task-irrelevant, but stable allocentric cues by increasing the width of the table (300 cm) and monitor (200 cm) so that they clearly exceeded the field of view that was covered by the head mounted display ($\sim 110^\circ$). We recruited 18 participants from which we had to exclude two participants because of technical problems during the experiment. From the remaining 16 participants (13 women; $M_{\text{age}} = 22.81$ years; right-handed, EHI: $M = 77.76$, $SD = 17.57$) we collected a total of 2592 trials from which we removed $\sim 3\%$ because movement onsets occurred before the start signal, and $< 1\%$ due to measurement errors. After aggregating the data, we removed $\sim 3\%$ of the aggregated data due to our reaching endpoint deviation criterion (2 *SD*).

Changes in Experiment 3

In Experiment 3, we controlled for variability in presentation depth of the objects between conditions by placing objects on shelves (Figure 1B, third column). In Visual Space, the shelves were placed in front of the participant such that the distance between the table edge and the edge of a shelf facing the participant was ~ 50 cm. In the Pictorial Space, the same shelves were shown on the screen of the virtual monitor. Each shelf was 500-cm wide (to ensure that the edges fall outside of the visual field), 1.3-cm thick, and 20-cm deep. The first shelf was placed ~ 7.5 cm above the table surface. The next two shelves at ~ 24.5 cm and ~ 41.5 cm height from table surface, respectively. To keep reaching trajectories fairly similar, we refrained from using the “Pictorial Small” condition here. Each of the six targets, that is, each arrangement was presented three times (baseline [no shift], shift left, shift right) in both display conditions resulting in $6 \times 3 \times 2 = 36$ trials. We measured each combination three times so that each participant completed 108 trials.

We recruited 19 participants from which we had to exclude one participant because of difficulties in

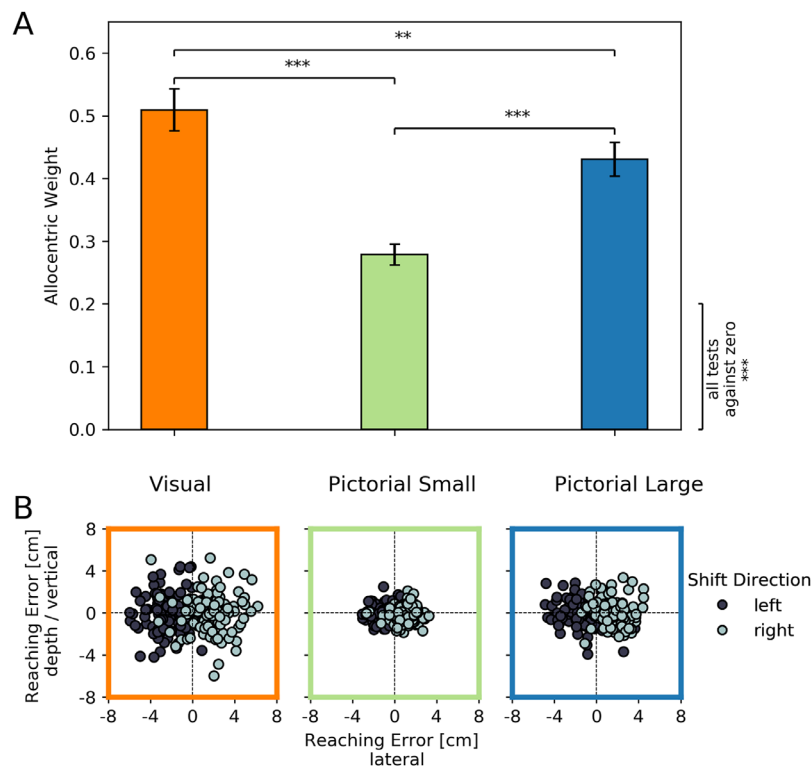


Figure 2. Results of Experiment 1 for the experimental conditions. (A) Allocentric weights including indicators for pairwise comparisons and error bars representing the standard error of the mean. (B) Reaching errors for the lateral and depth/vertical component for conditions in which table objects were shifted either to the left or to the right. $**p < 0.01$, $***p < 0.001$.

stereoscopic vision. From 18 participants (10 women; $M_{\text{age}} = 24.89$ years; right-handed, EHI: $M = 73.41$, $SD = 29.86$) we collected a total of $18 \times 108 = 1944$ trials. After data collection, in the postexperiment interview, one participant reported to have become aware of the objects' shifts. We removed the participant's trials leaving 1836 trials from which we removed $\sim 2.5\%$ because of movement onsets that occurred before the start signal. After aggregating the data, we removed $\sim 1\%$ of the aggregated data due to our reaching endpoint deviation criterion ($2SD$). We calculated the allocentric weights similar to the previous two experiments and compared the two conditions by means of a paired sample t -test.

Results

Experiment 1: Comparison of visual versus pictorial space

In Experiment 1, we asked if the strength of allocentric coding of reaching targets depends on whether objects are presented on a computer

screen in pictorial space or in the visual space of the observer. The results of Experiment 1 are depicted in Figure 2. There was a systematic effect of object shift with reaching endpoints deviating in the direction of object shift (Figure 2B), and allocentric weights significantly higher than when no reaching error would have occurred (all tests against zero: $p < 0.001$). The results show that there is a significant effect of *presentation mode*, $F_{2,34} = 53.125$, $p < 0.001$, $\eta_p^2 = 0.758$, with highest allocentric weights in the Visual Space condition ($M = 0.509$, $SD = 0.142$), intermediate weights in the Pictorial Large ($M = 0.431$, $SD = 0.114$), and the lowest weights in the Pictorial Small condition ($M = 0.279$, $SD = 0.071$). All pairwise comparisons were significant (Visual vs. Pictorial Small: $t(17) = 9.337$, $p < 0.001$, $d_z = 2.201$; Visual vs. Pictorial Large: $t(17) = 3.473$, $p = 0.003$, $d_z = 0.818$; Pictorial Small vs. Pictorial Large: $t(17) = -7.327$, $p < 0.001$, $d_z = 1.727$). Our results demonstrate that participants integrate allocentric information in both pictorial and visual space, even when we control for the retinal size of objects. In contrast to our hypothesis, we observed higher allocentric weights for spatial coding in Visual compared with Pictorial space.

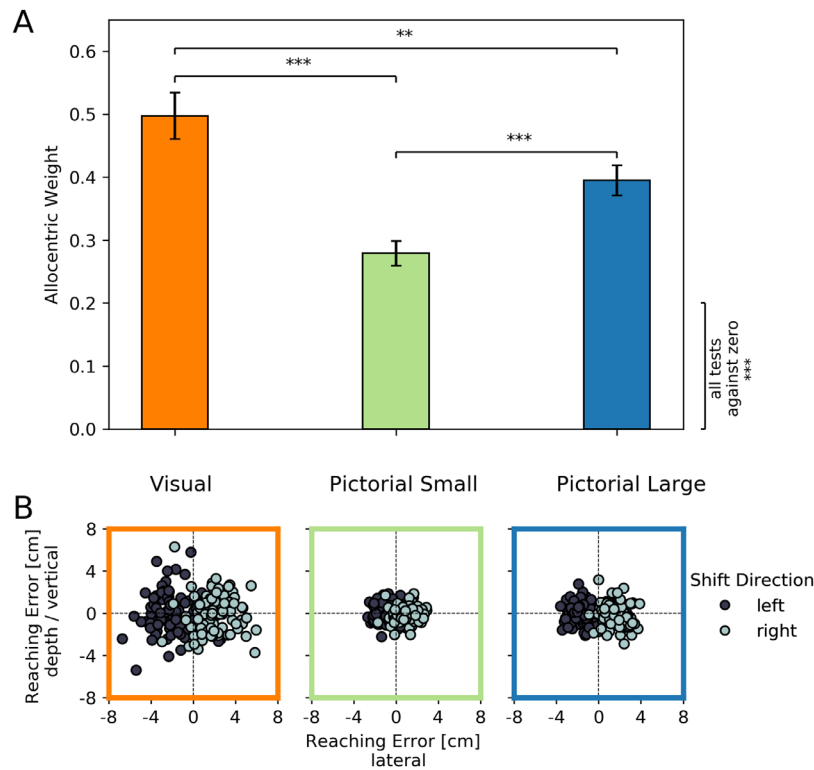


Figure 3. Results of Experiment 2 for all three experimental conditions. (A) Allocentric weights including indicators for pairwise comparisons and error bars representing the standard error of the mean. (B) Reaching errors for the lateral and depth/vertical component for conditions in which table objects were shifted either to the left or to the right. $**p < 0.01$, $***p < 0.001$.

Experiment 2: Allocentric cue control

The results of Experiment 1 show that we encode reach targets relative to other task-relevant objects (among other reference frames), which is reflected in the allocentric weights. Surprisingly, we seem to do so preferably in the Visual Space condition for which we found higher allocentric weights compared with the two Pictorial Space conditions. This suggests that, in the Pictorial Space conditions, participants represented target objects more strongly relative to other entities than the task-relevant objects. Such entities could be either the observer themselves, that is, egocentric reference frame, or other more stable task-irrelevant allocentric cues, for example, the frame of the monitor or the left and right table edges. Given the ill-defined location of the observer in pictorial space, we believe that a stronger reliance on egocentric information in the pictorial space seems to be less likely. Therefore in Experiment 2 we controlled for other potential allocentric cues participants could have used to represent the task-relevant objects. To this end, we substantially increased the width of the table and the monitor to reduce the possibility that their vertical edges are used as allocentric cues. If the vertical

monitor edges, as one of the task-irrelevant allocentric cues, were responsible for differences in spatial coding between Visual Space and Pictorial Space, we would expect them to be less pronounced. In contrast, if these allocentric cues did not play an important role for the encoding of the target position, we would expect to replicate the results of Experiment 1.

The results of Experiment 2 are depicted in Figure 3. We found that reaching endpoints systematically deviated in the direction of object shift (Figure 3B), with allocentric weights being significantly higher than when no reaching error would have occurred (all tests against zero: $p < 0.001$). Similar to the results in Experiment 1, we found a significant effect of *presentation mode*, $F_{2, 30} = 40.863$, $p < 0.001$, $\eta_p^2 = 0.731$, with highest allocentric weights in the Visual Space condition ($M = 0.497$, $SD = 0.147$), followed by the Pictorial Large ($M = 0.395$, $SD = 0.096$), and then the Pictorial Small condition ($M = 0.279$, $SD = 0.078$). All pairwise comparisons were significant (Visual vs. Pictorial Small: $t(15) = 7.927$, $p < 0.001$, $d_z = 1.982$; Visual vs. Pictorial Large: $t(15) = 3.767$, $p = 0.002$, $d_z = 0.942$; Pictorial Small vs. Pictorial Large: $t(15) = -7.327$, $p < 0.001$, $d_z = 1.832$). Our results support the findings of

Experiment 1 suggesting that participants make more use of allocentric information in visual compared with pictorial space.

Experiment 3: Control for variability in depth

The results of the first two experiments suggest that participants use an allocentric reference frame to encode task-relevant objects in both visual and pictorial space. Our experimental design was motivated by previous experiments: in some of them, participants were presented with images of the breakfast scene on a monitor, whereas in others, they viewed the breakfast scene in virtual reality with objects placed in front of them on a table. Our comparison of both presentation modes is therefore limited by differences in relative disparities. In Visual Space, objects were located at different distances from the observer eliciting relative disparities. This was not the case in Pictorial Space in which objects were all lying on a flat surface. Another limiting factor is that the two presentation modes also required different reaching movements, owing to differences in the vertical and depth location of the presented objects. To address these possible confounds, we created a third experiment in which we placed the six objects on shelves in front of the participants. By doing so, we keep the locations of objects on the horizontal and vertical plane comparable between Visual Space and Pictorial Space. Because one of our goals was to obtain comparable reaching kinematics, we refrained from a Pictorial Small condition and just compared the Pictorial Large and the Visual Space condition that are similar in retinal object size.

The results of Experiment 3 are depicted in Figure 4. We found again that reaching endpoints systematically deviated in the direction of object shift (Figure 4B), with allocentric weights being significantly higher than when no objects were shifted (both tests against zero: $p < 0.001$). We further found the allocentric weights in the Visual Space condition ($M = 0.519$, $SD = 0.146$) being significantly higher than in the Pictorial Large condition ($M = 0.449$, $SD = 0.136$); $t(16) = 3.077$, $p = 0.007$, $d_z = 0.746$. These results show that the findings from Experiment 1 were not owing to differences in presentation depth of objects or differences in reaching trajectories between the Visual Space and Pictorial Space condition.

Discussion

In a series of three experiments, we found that the brain computes spatial relations of task-relevant objects depending on whether they are presented in the pictorial space of a computer screen or in the visual

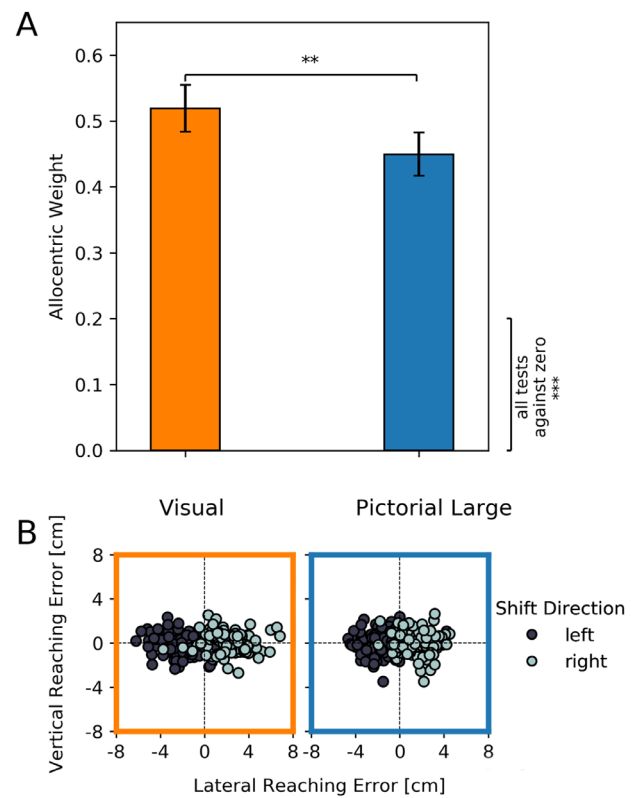


Figure 4. Results of Experiment 3 for both presentation modes. (A) Allocentric weights including indicators for a comparison of means and error bars representing the standard error of the mean. (B) Reaching errors for the lateral and vertical component for conditions in which the objects on the shelves were shifted either to the left or to the right. $**p < 0.01$.

space that the observer inhabits. In all conditions, we found significant allocentric weights, indicating that participants generally encoded task-relevant objects relative to each other. We thus replicated the original findings on the use of allocentric information for reaching with quite comparable weights ranging from approximately 30% to 50% (e.g., Fiehler et al., 2014; Klinghammer et al., 2016). Contrary to what we expected, participants encoded target objects relative to other task-relevant objects preferably when the objects were presented in visual space. This was reflected in higher allocentric weights in the Visual Space compared with both Pictorial Space conditions. In the second experiment, we wanted to ensure that differences in allocentric coding were not due to additional task-irrelevant but stable allocentric cues, namely the vertical edges of the monitor that might have served as a reliable landmark. In the third experiment, we controlled for variability in depth and reaching kinematics between the presentation modes by presenting objects on shelves, that is, coplanar with the frontoparallel plane, at comparable distance in both presentation modes. In the control experiments, we replicated the results

of the first experiment with remarkably similar values. Overall, our results suggest that the brain computes spatial relations of task-relevant objects depending on the space they belong to.

We expected lower allocentric weights in the Visual Space compared with the Pictorial Space conditions. This hypothesis was built on the fact that, by definition, the observer has a defined location, and therefore valid information about the distance and direction of objects in visual space, but not in pictorial space. We therefore expected allocentric information to be more reliable in pictorial than visual space. To explain this finding, we could try to think of ways how distinct features of the Pictorial Space conditions could have influenced our results. For example, the pictorial space is accessed through a pictorial plane that is most of the time facing the observer, for example, a computer monitor or a television screen. It is true that the pictorial plane is part of the visual space, and as such the argument of potential movements that make plane-to-observer (egocentric) coding less reliable is applicable here as well. Nevertheless, the prototypical observer position relative to the pictorial plane might have led to a change in the weighting of allocentric information. This hypothesis needs to be tested to understand if such prototypical observer positions exist and if they actually impact spatial coding.

One possible limitation of our study that might have decreased allocentric coding in the pictorial space is the presence of the monitor stand. The monitor stand was only visible in the Pictorial Space conditions, but not in the Visual Space condition. The stand may have provided a stable, task-irrelevant landmark for target encoding leading to lower allocentric weights in the Pictorial than the Visual Space conditions. However, such stable, task-irrelevant landmarks (the vertical edges of the monitor) did not influence allocentric coding in Experiment 1, as we observed similar allocentric weights irrespective of whether these landmarks were available (Experiment 1) or not (Experiments 2 and 3). This indicates that participants do not heavily rely on task-irrelevant allocentric cues when coding object locations; a finding that is in line with previous work from Klinghammer et al. (2015) using a similar object shift paradigm. They found that shifting task-irrelevant objects did not affect reaching endpoints, even if these shifts caused substantial changes in the scene. Another limitation is the lack of active depth cues. In visual space (i.e., real or virtual environments), we constantly use unique information, such as motion parallax, as a reliable depth cue to understand the geometry of the environment surrounding us. Any actual head movement would add just more information (depth cues) allowing us to create a more stable representation of the scene. In a pictorial space such information is absent. One could claim that our visual space in this study (virtual environment) does not completely

replicate the visual space we have in real-life because we limited participants' head movement by a chin rest. A chin rest, however, does not fully restrict head motion (as opposed to a bite-bar). Recent findings suggests that subtle head movements are already sufficient to influence depth judgments through motion parallax (de la Malla, Buiteman, Otters, Smeets, & Brenner, 2016). Further experiments should compare and clarify the effect of active depth cues, such as motion parallax, on computations of spatial representations.

In the first two experiments, we examined the effect of object size by adjusting the objects' retinal size in pictorial space to be similar (Pictorial Large) or dissimilar (Pictorial Small) to the objects' retinal size presented in visual space. This comparison resulted in higher allocentric weights when the objects' size in pictorial space matched the objects' size in visual space. This advantage for the "real-life" presentation of stimulus size might be explained by stronger visual affordances of objects or a higher potential for interaction (Gibson, 1979; Symes, Ellis, & Tucker, 2007). For example, an apple of 5° visual angle is more likely to be within one's reach space, and therefore more relevant for potential interactions as opposed to an apple of 1° visual angle. This hypothesis is in line with findings that show that humans rely on the retinal size of objects when estimating the distance between an object and an observer (Sousa, Brenner, & Smeets, 2011; Sousa, Smeets, & Brenner, 2013). In addition, object size has been shown to act as an effective depth cue when spatially representing objects in virtual reality (Klinghammer et al., 2016). In our third experiment, we refrained from this comparison but rather kept the size constant to improve comparability between both presentation modes.

At this point we can only speculate about neurophysiological accounts for our findings. One such account lies in the well-known concept of the two visual streams with a ventral and a dorsal stream that process visual information for perception relying on allocentric representations, and for action relying on egocentric representations, respectively (Goodale & Milner, 1992; Whitwell, Milner, & Goodale, 2014). These results were supported by studies showing that memory-guided movements processed by the ventral stream are susceptible to visual illusions, whereas real-time movements processed by the dorsal stream are unaffected (Hu & Goodale, 2000; Westwood & Goodale, 2003). In contrast, recent evidence suggests that the dorsal pathway also contributes to allocentric coding, and explains actions that are susceptible to visual illusions as well (Adam, Bovend'Eerd, Schuhmann, & Sack, 2016; de la Malla, Brenner, de Haan, & Smeets, 2019; Freud, Plaut, & Behrmann, 2016; Kravitz, Saleem, Baker, & Mishkin, 2011; Medendorp, de Brouwer, & Smeets, 2018). Based on these neurophysiological

accounts, the vision-for-action/dorsal pathway could become more important in spaces that are relevant for our actions (visual space), and in turn increases the use of task-relevant objects for allocentric coding.

Conclusions

Our findings allow a first insight into the processing of visual information linked to the space we are part of (visual space) and the space we can just be aware of (pictorial space) and suggests that the brain distinguishes between these two spaces. If the distinction between visual and the pictorial space matters, classic findings should be reinvestigated, thus allowing a more valid interpretation. Before doing so, we need to better understand if, for example, sensory (e.g., lack of motion parallax) and behavioral (e.g., lack of interaction) differences actually influence spatial coding, and whether they mediate computational differences between the visual space and the pictorial space.

Keywords: allocentric, egocentric, spatial representation, virtual reality, vision for action

Acknowledgments

The authors thank Jan Koenderink and Andrea van Doorn for helpful discussions, and Munyke Stelter for her help in data collection. This work was supported by the International Research Training Group (IRTG) 1901 “The brain in action” funded by the German Research Foundation (DFG), the DFG FI 1567/6-1 TAO (The active observer), and the VISTA Distinguished Visiting Scholar Program (York University, Toronto) to KF. NFT received funding from the NSERC Discovery Grant and the CFREF VISTA program. HK received a dissertation fellowship of the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes). Result data of all three experiments can be found at <https://doi.org/10.5281/zenodo.3723425>.

Commercial relationships: none.

Corresponding author: Harun Karimpur.

Email: harun.karimpur@psychol.uni-giessen.de.

Address: Experimental Psychology, Justus Liebig University, Giessen, Germany.

Footnote

¹ Throughout the manuscript, we capitalize the presentation modes, such as “Visual Space,” when we refer to experimental conditions.

References

- Adam, J. J., Bovend'Eerdt, T. J. H., Schuhmann, T., & Sack, A. T. (2016). Allocentric coding in ventral and dorsal routes during real-time reaching: Evidence from imaging-guided multi-site brain stimulation. *Behavioural Brain Research*, *300*, 143–149, <https://doi.org/10.1016/j.bbr.2015.12.018>.
- Arbib, M. A. (1991). Interaction of multiple representations of space in the brain. In J. Paillard (Ed.), *Brain and space* (pp. 379–403). New York: Oxford University Press.
- 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, <https://doi.org/10.1152/jn.01008.2009>.
- Camors, D., Jouffrais, C., Cottureau, B. R., & Durand, J. B. (2015). Allocentric coding: Spatial range and combination rules. *Vision Research*, *109*, 87–98, <https://doi.org/10.1016/j.visres.2015.02.018>.
- Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, *20*(1), 15–24, [https://doi.org/10.1016/S0896-6273\(00\)80429-8](https://doi.org/10.1016/S0896-6273(00)80429-8).
- Crawford, J. D., Henriques, D. Y. P., & Medendorp, W. P. (2011). Three-dimensional transformations for goal-directed action. *Annual Review of Neuroscience*, *34*(1), 309–331, <https://doi.org/10.1146/annurev-neuro-061010-113749>.
- de la Malla, C., Brenner, E., de Haan, E. H. F., & Smeets, J. B. J. (2019). A visual illusion that influences perception and action through the dorsal pathway. *Communications Biology*, *2*(1), 38, <https://doi.org/10.1038/s42003-019-0293-x>.
- de la Malla, C., Buiteman, S., Otters, W., Smeets, J. B. J., & Brenner, E. (2016). How various aspects of motion parallax influence distance judgments, even when we think we are standing still. *Journal of Vision*, *16*(9):8, 1–14, <https://doi.org/10.1167/16.9.8>.
- Diedrichsen, J., Werner, S., Schmidt, T., & Trommershäuser, J. (2004). Immediate spatial distortions of pointing movements induced by visual landmarks. *Perception and Psychophysics*, *66*(1), 89–103, <https://doi.org/10.3758/BF03194864>.
- 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, <https://doi.org/10.3389/fnhum.2014.00636>.
- Freud, E., Macdonald, S. N., Chen, J., Quinlan, D. J., Goodale, M. A., & Culham, J. C. (2018).

- Getting a grip on reality: Grasping movements directed to real objects and images rely on dissociable neural representations. *Cortex*, 98, 34–48.
- Freud, E., Plaut, D. C., & Behrmann, M. (2016). ‘What’ is happening in the dorsal visual pathway. *Trends in Cognitive Sciences*, 20(10), 773–784, <https://doi.org/10.1016/j.tics.2016.08.003>.
- Gerhard, T. M., Culham, J. C., & Schwarzer, G. (2016). Distinct visual processing of real objects and pictures of those objects in 7- to 9-month-old infants. *Frontiers in Psychology*, 7, 1–9, <https://doi.org/10.3389/fpsyg.2016.00827>.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Goldstein, E. B. (1987). Geometry or not geometry? Perceived orientation and spatial layout in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 256–266, <https://doi.org/10.1037/0096-1523.14.2.312>.
- Gomez, M. A., & Snow, J. C. (2017). Action properties of object images facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1115–1124, <https://doi.org/10.1037/xhp0000390>.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Cognitive Sciences*, 15(1), 20–25, [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8).
- Hu, Y., & Goodale, M. A. (2000). Grasping after a delay shifts size-scaling from absolute to relative Metrics. *Journal of Cognitive Neuroscience*, 12(5), 856–868.
- Karimpur, H., Morgenstern, Y., & Fiehler, K. (2019). Facilitation of allocentric coding by virtue of object-semantics. *Scientific Reports*, 9, 1–9, <https://doi.org/10.1038/s41598-019-42735-4>.
- Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition-An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 1–17). Berlin: Springer, https://doi.org/10.1007/3-540-69342-4_1.
- 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>.
- 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, <https://doi.org/10.3389/fnins.2017.00204>.
- 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, <https://doi.org/10.1016/j.visres.2016.10.004>.
- Koenderink, J. J., & van Doorn, A. (2008). The structure of visual spaces. *Journal of Mathematical Imaging and Vision*, 31, 171–187, <https://doi.org/10.1007/s10851-008-0076-3>.
- Koenderink, J. J., & van Doorn, A. (2012). Gauge fields in pictorial space. *SIAM Journal on Imaging Sciences*, 5(4), 1213–1233, <https://doi.org/10.1137/120861151>.
- Koenderink, J. J., & van Doorn, A. J. (2003). Pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking into pictures: An interdisciplinary approach to pictorial space* (pp. 239–299). Cambridge, MA: MIT Press.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230, <https://doi.org/10.1038/nrn3008>.
- Krigolson, O., & Heath, M. (2004). Background visual cues and memory-guided reaching. *Human Movement Science*, 23(6), 861–877, <https://doi.org/10.1016/j.humov.2004.10.011>.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12, <https://doi.org/10.3389/fpsyg.2013.00863>.
- 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>.
- Marini, F., Breeding, K. A., & Snow, J. C. (2019). Distinct visuo-motor brain dynamics for real-world objects versus planar images. *NeuroImage*, 195, 232–242.
- Medendorp, W. P. (2011). Spatial constancy mechanisms in motor control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 476–491, <https://doi.org/10.1098/rstb.2010.0089>.
- Medendorp, W. P., de Brouwer, A. J., & Smeets, J. B. J. (2018). Dynamic representations of visual space for perception and action. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 98, 194–202, <https://doi.org/10.1016/j.cortex.2016.11.013>.
- 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, <https://doi.org/10.1007/s00221-005-0055-5>.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113, [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4).
- Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., . . . Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, 1(130), 1–10, <https://doi.org/10.1038/srep00130>.
- Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, 8, 1–11, <https://doi.org/10.3389/fnhum.2014.00837>.
- Sousa, R., Brenner, E., & Smeets, J. B. J. (2011). Judging an unfamiliar object's distance from its retinal image size. *Journal of Vision*, 11(9):10, 1–6, <https://doi.org/10.1167/11.9.10.Introduction>.
- Sousa, R., Smeets, J. B. J., & Brenner, E. (2013). The influence of previously seen objects' sizes in distance judgments. *Journal of Vision*, 13(2):2, 1–8, <https://doi.org/10.1167/13.2.2.doi>.
- Symes, E., Ellis, R., & Tucker, M. (2007). Visual object affordances: Object orientation. *Acta Psychologica*, 124(2), 238–255, <https://doi.org/10.1016/j.actpsy.2006.03.005>.
- Taghizadeh, B., & Gail, A. (2014). Spatial task context makes short-latency reaches prone to induced Roelofs illusion. *Frontiers in Human Neuroscience*, 8(673), 1–13, <https://doi.org/10.3389/fnhum.2014.00673>.
- Troje, N. F. (2019). Reality check. *Perception*, 48(10), 1–6.
- Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, 8(10), 1401–1410, <https://doi.org/10.1038/nn1553>.
- Westwood, D. A., & Goodale, M. A. (2003). Perceptual illusion and the real-time control of action. *Spatial Vision*, 16, 243–254, <https://doi.org/10.1163/156856803322467518>.
- Whitwell, R. L., Milner, A. D., & Goodale, M. A. (2014). The two visual systems hypothesis: New challenges and insights from visual form agnostic patient DF. *Frontiers in Neurology*, 5(255), 1–8, <https://doi.org/10.3389/fneur.2014.00255>.
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194, <https://doi.org/10.1001/jama.2013.281053>.