Allocentric and egocentric cues constitute an internal reference frame for real-world visual search

Yan Chen[a, b, †] and Zhe-Xin Xu[a, c, †]

[a]Shanghai Key Laboratory of Brain Functional Genomics, Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

[b]Department of Psychology, University at Buffalo, Buffalo, NY, USA

[c]Department of Neurobiology, Harvard Medical School, Boston, MA, USA

[†]These authors contributed equally.

Abstract

Visual search in natural environments involves numerous objects, each composed of countless features. Despite this complexity, our brain efficiently locates targets. Here, we propose that an internal reference frame facilitates real-world visual search. Objects in natural scenes often appear in orientations perceived as upright, enabling quicker recognition. However, how object orientation influences real-world visual search remains unclear. Moreover, the contributions of multiple reference cues—egocentric, visual context, and gravitational—on visual search are not well understood. To answer these questions, we designed a visual search task in virtual reality. Our results revealed an orientation effect that was independent of set size, suggesting the involvement of internal reference frame transformation. By independently rotating the virtual scene and participants in a flight simulator, we found that allocentric cues significantly altered search performance. These findings provide fundamental insights into the efficiency of real-world visual search and its connection to high-level cognitive processes.

Keywords: real-world visual search, virtual reality, allocentric reference frames, mental rotation

Introduction

Imagine navigating a busy mall looking for your friend amidst the crowd, or spotting a bottle of juice among stacks of products in a grocery store. These seemingly mundane moments illustrate the remarkable efficiency of visual search, where the nervous system sifts through countless colors, shapes, and motions to find a specific target. How does our brain manage to find that needle in a haystack? What cues guide visual search in natural environments?

While decades of research have uncovered the elements of visual search, such as parallel and serial search[1-5], traditional theories are limited to simplified, controlled

settings and fail to explain search performance involving natural scenes. For example, feature integration theory[1,2] predicts that visual search in the real world should be agonizingly slow, as it requires a step-by-step integration of numerous features. In practice, however, our brain performs these tasks with surprising speed and accuracy[5,6].

Recent studies have revealed several factors contributing to this striking efficiency of real-world visual search, including top-down attention, stimulus saliency, semantic context, prior history, and environmental statistics[7-13]. These factors establish a priority map that guides attention to specific features or spatial locations[9,10]. Another key element of natural scenes involves objects frequently occurring in consistent orientations[14-16], which may guide visual search. While studies have established that humans recognize upright stimuli more quickly and accurately[14,16-24], it is unclear how object orientation influences visual search.

Moreover, our brain represents the world in multiple reference frames[25-35]. For example, when tilting your head to peer through an aisle in a grocery store, the bottle of juice might appear tilted in your eyes. Nevertheless, the brain correctly identifies the bottle by integrating various reference cues, including how objects are oriented relative to the body (egocentric), the surrounding environment (visual context), and gravity[33,36-39]. To date, the role of different reference frames in real-world visual search remains mysterious.

To address this major knowledge gap, we design a real-world visual search task in virtual reality. We demonstrate a shorter response time (RT) and higher accuracy for upright objects than those oriented horizontally. This orientation dependency is independent of set size, indicating a transformation of an internal reference frame rather than individual object rotation. By pairing a flight simulator with a head-mounted display, we disentangle the contributions of different reference cues that are normally overlapped. We show that orientation dependency is significantly altered by visual context and gravitational cues, demonstrating the involvement of multisensory signals and high-level

cognitive processes. This study provides a systematic examination of how object orientation influences visual search in naturalistic environments, offering insights into visual processing in the natural world.

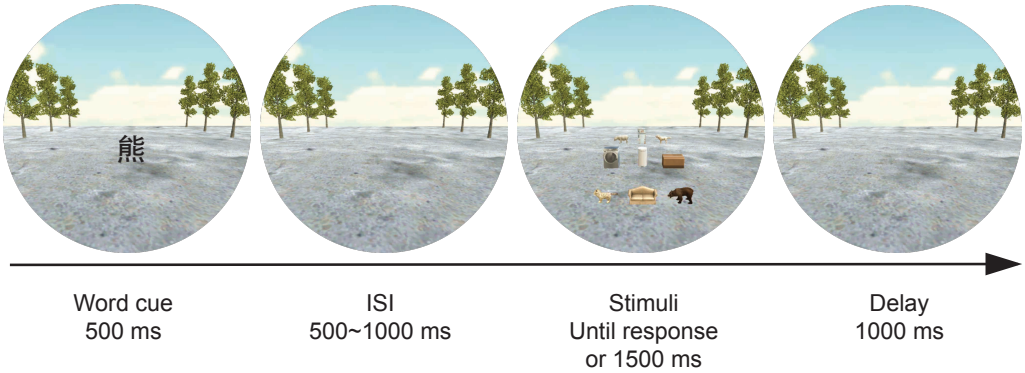

Figure 1 Procedure and stimuli. At the beginning of each trial, a word cue was presented at the center of the display (e.g., “Bear”), indicating the present trial's search target. After a variable inter-stimuli interval (ISI), objects appeared for 1500 ms or until a response was made by the participant. The next trial started after a 1000-ms delay.

Results

Orientation dependency in real-world visual search

We started by asking how object orientation influences the performance of real-world visual search. We designed a psychophysical task in a virtual reality system that consisted of a head-mounted display and a flight simulator (see Methods). Objects were presented in a natural scene in the virtual environment, and participants were asked to report whether a cued target was present (Figure 1). Eight experimental conditions were introduced, including the combinations between two set sizes (4 and 9), two object orientations (0°, upright; 90°, horizontal), and target presence (present or absent).

All participants achieved accuracies well above chance, ranging from 81.4% to 95.6%, with a mean of 89.4%. The hit rate ranged from 71.2% to 94.7%, with a mean of 84.4%, and the false alarm rate ranged from 3.2% to 10.1%, with a mean of 5.6%. Discriminability ranged from 1.89 to 3.43, with a mean of 2.65.

We used a linear mixed-effects model to quantify the effects of various factors on RT

while accounting for variability across participants and object categories. The model included object orientation, target presence, set size, and interactions between them as fixed effects and subject identity and object category as random effects (see Methods). We found a significant effect of object orientation on normalized RT, with longer RT for upright objects than for horizontal ones (Figure 2A; estimate=0.002, 95%CI=[0.00055, 0.0034], $t(10029)=2.7$, $p=0.00695$). Participants also showed higher accuracy for upright objects than for horizontal ones (Figure 2B; estimate=-5.02×10^{-4}, 95%CI=[-9.69×10^{-4}, -3.63×10^{-5}], $t(11146)=-2.11$, $p=0.0346$).

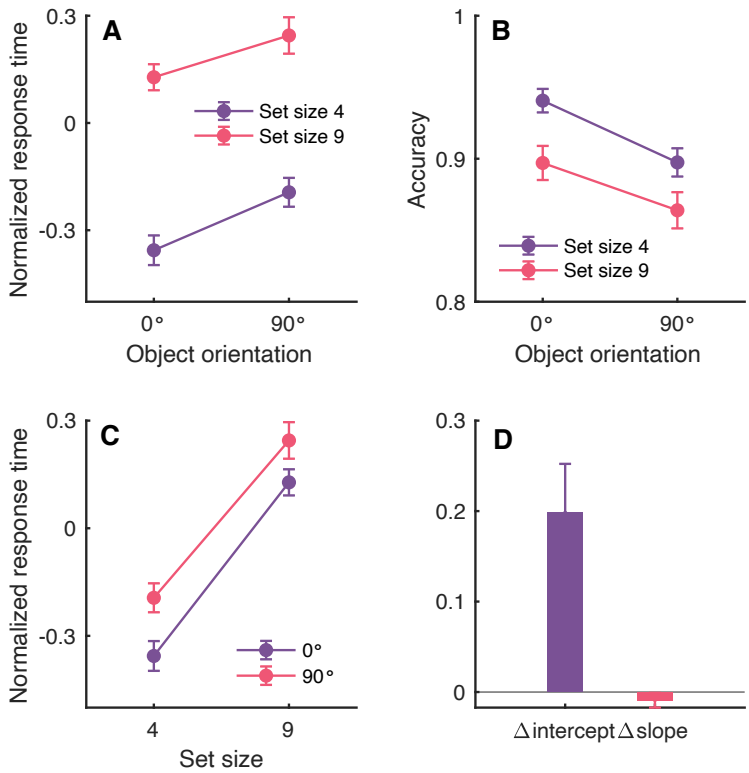

Figure 2 Results for Experiment 1. A, Normalized response time was longer for horizontally oriented objects (90°) than for upright objects (0°) for set sizes 4 (purple line) and 9 (red line). B, The accuracy was higher when objects were upright than when they were horizontal. Format as in A. C, Normalized response time increased as a function of set size for upright (purple line) and horizontal objects (red line). D, Difference in intercept and slope of the response time–set size function (lines shown in C) between upright and horizontal objects. Error bars represent S.E.M.

These results demonstrate that real-world visual search is faster and more accurate

for upright objects than for horizontal objects, suggesting an orientation dependency in this task. In addition, we found that normalized RT scaled with set size (Figure 2C; estimate=0.129, 95%CI=[0.116,0.142], t(10029)=19.1, $p=1.26\times10^{-79}$), which suggests that a serial search was involved.

Internal reference frame facilitates visual search

What cognitive process might give rise to this orientation dependency in visual search? There are two possibilities: individual object rotation and internal reference frame transformation. As discussed above, the effect of set size on RT indicates a serial search during which individual objects are processed sequentially. If an observer mentally rotates each object during this serial process, we expect set size to have a greater effect on RT for tilted objects than for upright objects. In other words, for each additional object in the scene, additional process time is needed to mentally rotate that object, resulting in a steeper slope of the RT–set size function for tilted than for upright objects. Alternatively, the observer may rotate an internal reference frame to align it with objects and maintain its orientation throughout the serial process. Because an internal reference frame transformation happens before serial search and does not depend on the number of objects present, we expect the effect of object orientation to be independent of set size.

Our linear mixed-effects model showed that the interaction between object orientation and set size was not significant (estimate=-1.30×10^{-4}, 95%CI=[-3.39×10^{-4}, 7.91×10^{-5}], t(10029)=-1.22, p=0.223). Furthermore, we observed a significant difference in the intercepts of the RT–set size functions for upright and horizontal objects (Figure 2D; m=-0.198, 95%CI=[-0.311, -0.086], t(19)=-3.68, p=0.00159, paired t-test), but not in the slopes (Figure 2D; m=0.009, 95%CI=[-0.008, 0.0261], t(19)=1.11, p=0.281). These findings suggest an internal reference frame transformation, rather than individual object rotation.

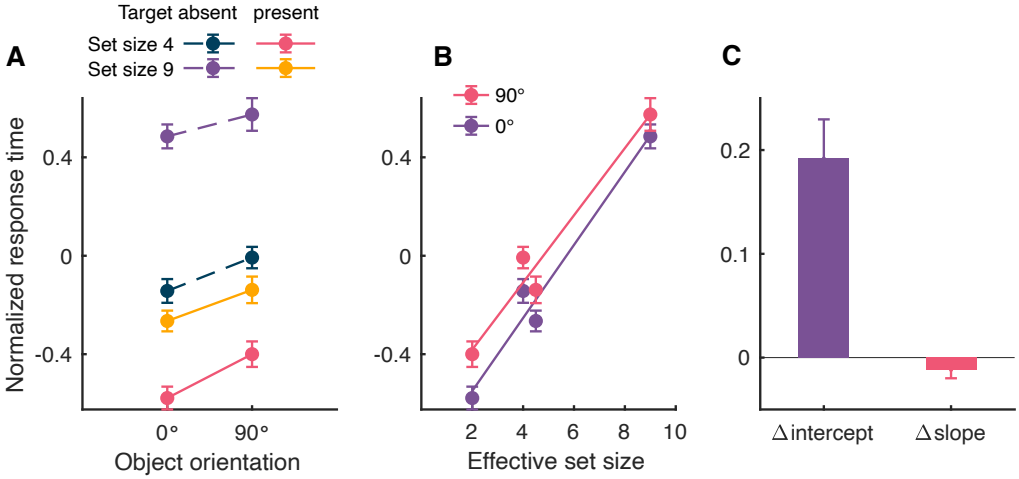

Figure 3 Results for target present and absent trials in Experiment 1. A, Longer normalized response time was found for both target absent (dashed lines) and present trials (solid lines) for set sizes 4 (dashed blue and solid red lines) and 9 (dashed purple and solid yellow lines). B, Response time varied as a function of effective set size for upright (purple) and horizontal objects (red). Lines represent linear fit to the data. C, Difference in intercept and slope of the response time– effective set size functions shown in B. Error bars indicate S.E.M.

We also observed a significant difference between target present and absent trials (Figure 3A; estimate=-0.160, 95%CI=[-0.291, -0.0283], $t(10029)=-2.383$, $p=0.0172$) and a significant interaction between target presence and set size (estimate=-0.0655, 95%CI=[-0.0846, -0.0464], $t(10029)=-6.730$, $p=1.793\times10^{-11}$). Interestingly, target presence did not interact with object orientation significantly (Figure 3A; estimate=3.85×10^{-4}, 95%CI=[-1.71×10^{-3}, 2.48×10^{-3}], $t(10029)=0.360$, $p=0.719$). To reconcile these observations, we introduced an effective set size to formalize the relationship between set size and target presence. We reasoned that in target-absent trials, an observer must process all objects in the scene before asserting the target is absent. Therefore, the effective set size is the same as the actual set size. On the other hand, in target-present trials, the observer can reach a decision and terminate the search process whenever the target is found. On average, only half of the objects would be processed before the target is identified. As a result, the effective set size in target-present trials is half the actual set size. We hypothesized that RT increases linearly with effective set size and that object orientation affects the intercept but not the slope of the

RT–effective set size function.

Indeed, we found a strong correlation between effective set size and normalized RT (Figure 3B; r=0.871, p=8.95×10^{-26} for upright objects, r=0.816, p=3.22×10^{-20} for tilted objects, Pearson's correlation). Importantly, the intercept, but not the slope, of the RT–effective set size function differs significantly between the two object orientations (Figure 3C; m=-0.192, 95%CI=[-0.271, -0.113], t(19)=-5.09, p=6.44×10^{-5} for intercepts; m=0.0119, 95%CI=[-0.00478, 0.0287], t(19)=1.49, p=0.151 for slopes, paired t-test), consistent with our findings of the RT–set size function.

Together, these results demonstrate that the effects of object orientation on RT were induced by internal reference frame transformation, which is independent of set size and stable for seconds within each trial.

Allocentric cues contribute to orientation dependency

The findings in the previous section suggest that an internal reference frame is transformed to facilitate visual search for tilted objects. What cues might determine this internal reference frame during a real-world search? Two general types of reference frames might be involved: egocentric and allocentric reference frames[25,32,33]. In egocentric reference frames, objects are represented relative to the retina, head, or body. In allocentric reference frames, visual context and gravitational cues may be used as references[15,28,35-37,39-41]. To quantify the effects of different reference frames, we manipulated the visual context and egocentric cues by leveraging a virtual reality system consisting of a flight simulator and a head-mounted display (Figure 4).

Four conditions were interleaved across blocks, including 1) a Baseline condition in which neither the visual context nor the flight simulator rotated, 2) a Visual context condition in which the visual scene was rotated by 90° (Figure 4B), 3) an Egocentric condition in which the flight simulator rotated by 90° (Figure 4A) while the visual context remained upright, and 4) a Gravitational condition, in which both the flight simulator and

visual context rotated by 90° in the same direction, such that gravitational cues differed from these two cues.

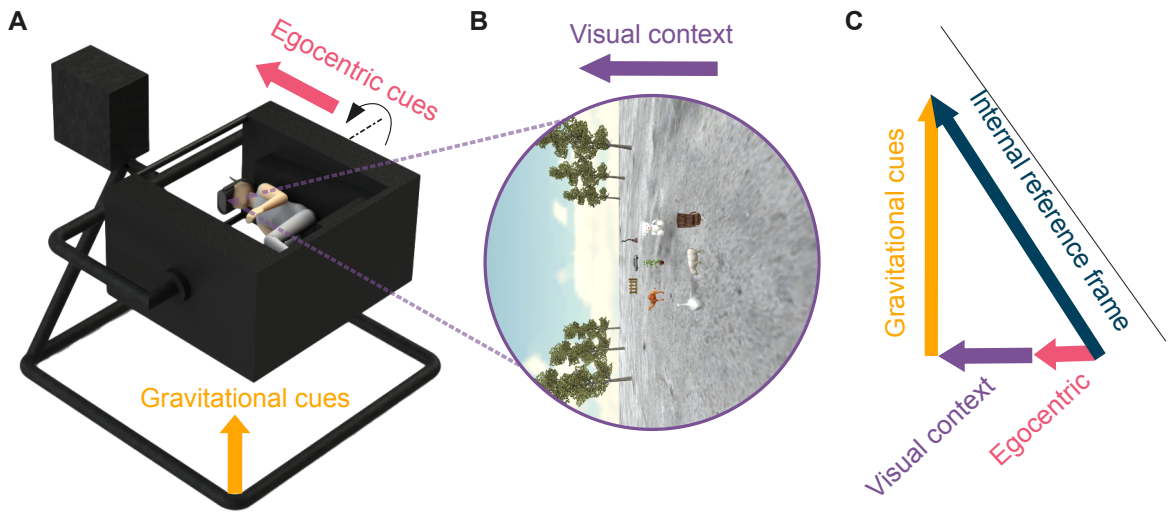

Figure 4 Apparatus and conditions in Experiment 2. A, Participants sat in a flight simulator, which allowed full body roll on a block-by-block basis, thus manipulating the egocentric reference cues. B, Visual context was manipulated independently by rotating the virtual environment. C, The internal reference frame can be modeled as a vector sum of visual context, egocentric, and gravitational cues. Note that while gravitational cues were not directly manipulated, they can be dissociated from the other two cues by simultaneously rotating the visual context and the participant's body.

Similar to Experiment 1, we found a significant effect of object orientation on normalized RT (Figure 5A; estimate=1.64×10^{-3}, 95%CI=[8.13×10^{-4}, 2.47×10^{-3}], $t(9865)=3.89$, $p=1.02\times10^{-4}$, linear mixed-effects model; see Methods). Importantly, we found significant interactions between object orientation and visual context (estimate=-3.11×10^{-5}, 95%CI=[-4.40×10^{-5}, -1.81×10^{-5}], $t(9865)=-4.69$, $p=2.73\times10^{-6}$), between object orientation and egocentric cues (estimate=-2.25×10^{-5}, 95%CI=[-3.54×10^{-5}, -9.50×10^{-6}], $t(9865)=-3.40$, $p=6.84\times10^{-4}$), and among object orientation, visual context, and egocentric cues (estimate=6.61×10^{-5}, 95%CI=[4.58×10^{-7}, 8.65×10^{-7}], $t(9865)=6.36$, $p=2.14\times10^{-10}$). These findings suggest that all three reference cues contribute to orientation dependency observed in our visual search task.

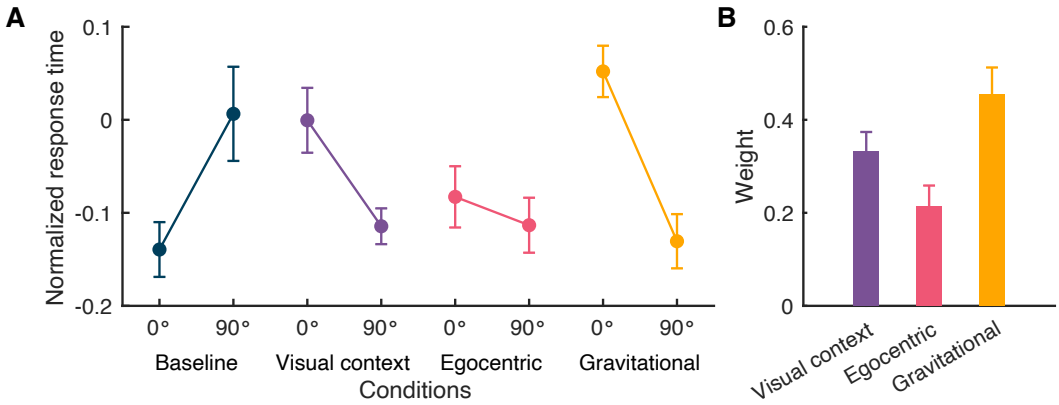

Figure 5 Results for Experiment 2. A, Normalized response time in each experimental condition. 0° and 90° represent upright and tilted objects, respectively. In each condition, object orientation was defined relative to the reference cues that were not under experimental manipulation (i.e., gravitational reference frame in the Baseline, Visual context, and Egocentric conditions; egocentric reference frame in the Gravitational condition). B, Weights for each reference frame inferred from changes in orientation dependency. Error bars represent S.E.M.

We computed the relative weights for three reference frames by comparing an orientation dependency index for each condition (Figure 4C and 5B; see Methods). The weights for all three references were significantly higher than zero (estimate=0.33, 95%CI=[0.23, 0.43], $t(27)=7.17$, $p=1.04\times10^{-7}$ for visual context; estimate=0.21, 95%CI=[0.12, 0.31], $t(27)=4.61$, $p=8.77\times10^{-5}$ for egocentric cues; estimate=0.45, 95%CI=[0.36, 0.55], $t(27)=9.80$, $p=2.18\times10^{-10}$ for gravitational cues; linear mixed-effects model). An analysis of variance revealed that the weights varied significantly across reference cues: $F(3, 27)=56.22$, $p=9.71\times10^{-12}$. Gravitational cues had the highest weight, visual context ranked second, and the weight for egocentric cues was the lowest (Figure 5B). Tukey's HSD test revealed that the weight for gravitational cues is significantly higher than that for egocentric cues (mean difference=0.24, 95%CI=[0.07, 0.41], $p=0.0047$), but the differences between other pairs of weights were not significant (mean difference=0.12, 95%CI=[-0.05, 0.29], $p=0.217$ between visual context and egocentric cues; mean difference=0.12, 95%CI=[-0.05, 0.29], $p=0.200$ between gravitational and visual context cues).

These results demonstrated a drastic change in orientation dependency when different reference frames were manipulated. This confirms our findings in Experiment 1

that an internal reference frame is involved during visual search for tilted objects. Furthermore, these results suggest that multiple ego- and allocentric reference cues constitute the internal reference frame, which may help to establish a stable representation of the world (see Discussion).

Discussion

In this study, we investigated the influences of object orientation and reference cues on visual search in natural scenes. We found that search performance was higher for upright objects compared to titled ones, demonstrating an orientation dependency during real-world search. Two possible mechanisms of orientation dependency were considered: internal reference frame transformation and individual object rotation. Further analysis of the relationship between RT and set size revealed that only the intercept, but not the slope, of the RT–set size function differs between upright and tilted objects. This finding rules out the possibility of individual object rotation, which predicts a change in the slope of the RT–set size relationship. Instead, our findings suggest that the brain employs an internal reference frame for visual processing, and this internal reference was rotated to match the orientation of objects before the search. By rotating the participants' bodies and the visual scene, we showed that multiple reference cues, including egocentric, visual context, and gravitational, contribute to the internal reference frame. Allocentric reference frames, namely visual context and gravitational cues, had greater influences on the orientation dependency than egocentric cues, suggesting the involvement of multisensory information and high-level cognitive processing.

What is the advantage of using an internal reference frame? Transforming an internal reference frame, rather than sequentially rotating each object, significantly reduces the computational complexity of visual search in the real world. By rotating the reference frame once before initiating a search, the process requires a fixed amount of cognitive resources that do not depend on set size. This allows the brain to process complex natural scenes in a more energy-efficient manner. In contrast, if objects are rotated

individually, each additional object would necessitate additional cognitive resources, making the process more resource-intensive as the visual scene becomes more complex. This study, along with other studies on reference frames[17,24-30,33,34,36,41], suggests that the nervous system adapts to environmental statistics and efficiently processes information based on these statistics. Our findings reveal that internal reference frame transformation might be another strategy the brain uses to improve visual search efficiency in real-world environments, complementing existing factors[5,6,9-13,40,42].

Visual information is primarily encoded in a retinotopic space in the early stages of visual processing[43-45]. However, a retinotopic representation is often unstable due to movements of the eyes, head, and body as we navigate the environment. To form a stable representation of the world, information about posture and self-motion—including efference copy of motor commands and sensory inputs from multiple modalities—is integrated with visual inputs to form an allocentric representation[25,26,28,30-32,34-37,39,41,46-54]. This transformation from egocentric to allocentric references may be initiated in V1[55-57], continued in higher-level visual areas[48,49,53], and propagated to the parietal cortex[26,31,32,34,35,46,47].

Our results show that real-world visual search relies heavily on gravitational cues, which may include vestibular signals and proprioception[35,37,41]. The greater reliance on these cues might be driven by the complexity of natural objects and scenes, which requires the engagement of high-level areas[9]. The visual system might rely more on retinotopic or egocentric reference frames for higher precision when processing simple visual features such as oriented bars and intersecting line segments[14,16,24,33,58]. An interesting future direction is to examine the relative weightings of different reference cues in tasks involving stimuli ranging from simple to complex. It is possible that the nervous system flexibly switches between multiple reference frames and adaptively chooses the optimal reference in each context[26].

In summary, this study demonstrates that humans use an internal reference frame in

real-world visual search. Using a virtual reality system, we show that multiple allocentric and egocentric reference cues shape this internal reference frame, providing a novel explanation for visual search efficiency in naturalistic environments. Our findings suggest that real-world visual search offers valuable insights into high-level cognitive processes that involve multiple sensory modalities.

Methods

Participants

Twenty participants (seven males and thirteen females, mean age 20.40, SD=0.49) with normal or corrected-to-normal vision participated in Experiment 1, and ten of them (five males and five females, mean age 20.90, SD=0.80) participated in Experiment 2. All subjects were new to psychophysical experiments and were unaware of the purpose of the study. Informed written consent was obtained from each participant before data collection. The study was approved by the Institutional Review Board at East China Normal University.

Apparatus

A head-mounted display (HTC Vive, first generation, HTC and Valve Corporation) was used for stimuli presentation. The mean luminance of the display was 61.15 cd/m^2, the resolution was 1080×1200 pixels, the refresh rate was 90 Hz, and the pixel size was ~0.09°×0.09°. The field of view of the head-mounted display was 111.5°. Visual stimuli were generated in Unity3D (Unity Technologies, version 5.6.2). Participants sat on a chair inside a custom flight simulator and were secured by a seatbelt and additional paddings around the head and body (Figure 4A). Participants held a wireless joystick to respond to the task. The flight simulator could be rotated along the roll axis at low speed.

Stimuli

The visual stimuli consisted of an outdoor scene with objects placed on the ground (Figure 1). A matrix of either four or nine objects was presented at the center of the screen in an area measuring 49°×53°. The size of each object was around 4.8°×4.8°. Objects included large animals and furniture with a canonical orientation. Their orientation was either 0° (upright) or 90° (horizontal), with all objects sharing the same orientation during each trial. Objects presented on each trial were randomly selected from 89 categories, each comprising two variations with different shapes and colors. All 3D models used in the experiments were acquired from the Unity Asset Store.

Experimental conditions

In Experiment 1, two set sizes, 4 and 9, were interleaved across blocks. Each block consisted of 72 trials. Two object orientations, horizontal and upright, were interleaved across trials. Target absent and present trials were interleaved and counterbalanced within each block. The flight simulator and visual context did not rotate throughout the experiment. Each participant completed two sessions with four blocks in each session.

In Experiment 2, Four reference frame conditions were introduced. 1) In the Baseline condition, the flight simulator and visual context remained upright relative to gravity. 2) In the Visual context condition, the visual scene was rotated 90° in either direction relative to gravity. 3) In the Egocentric condition, the flight simulator rotated 90° while the visual context remained upright. 4) In the Gravitational condition, both the flight simulator and visual context rotated 90° in the same direction such that only gravitational cues differed from the other two cues. These four conditions were interleaved across blocks. The set size was 9 throughout the experiment. Object orientation and target presence conditions were interleaved in the same way as in Experiment 1. Each participant completed two sessions with eight blocks in each session. Importantly, the head-mounted display covered the entire field of vision and did not move relative to the observer's head throughout the experiment, allowing complete control over the visual information

received by the participants.

Procedure

On each trial, a word cue was presented for 500 milliseconds, indicating the search target for the current trial (Figure 1). After a variable inter-stimuli interval (ISI), a set of objects was presented on the screen. The ISI was randomly selected from a set of discrete values ranging from 500 to 1000 milliseconds with a step of 100 milliseconds. Participants were asked to report whether the target was present by pressing one of the two buttons on the joystick. The objects disappeared after a response was made. If a response was not made within 1500 milliseconds, the objects would disappear, and participants were still forced to make a response. The next trial started 1000 milliseconds after a response was made.

Data preprocessing

For each participant, trials with RT greater than 4 median absolute deviations (MADs) from the median were removed from further analysis. MAD was used because it is more robust to outliers than standard deviation. On average, 2.67% of trials were removed. RTs were normalized by taking the Z score for each participant. Only correct trials were used for analyses of RTs.

Data analysis

Discriminability. The discriminability of each subject was calculated as: $d' = Z(hit) - Z(false\ alarm)$, where $Z(hit)$ was the Z score of hit rate and $Z(false\ alarm)$ was the Z score of the false alarm rate.

Effective set size. In target-absent trials, an observer must identify all objects before responding. Therefore, the effective set size would be the same as the actual set size. In contrast, when the target is present, only half of the objects, on average, need to be

processed before the target is identified. Therefore, the effective set size would be half of the actual set size in target-present trials.

RT–set size function. A linear function, $z(RT) = b_1 + b_2 x$, was fit to the relationship between (effective) set size x and the z score of RT, where b_1 and b_2 are the intercept and slope of the function, respectively. Fitting was done using the "\" operator in MATLAB (MathWorks, MA).

Linear mixed-effects models. Because of the nature of RT data, there are large variabilities across individuals and trials [59]. Traditional statistical analyses, such as the analysis of variance (ANOVA) on group means, do not consider these variabilities and often underestimate the true effects [59]. Here, we modeled RT using linear mixed-effects models that considered participants and object categories as random effects on the RT intercept and experimental conditions as fixed effects. These models were fit to trial-level data and could better account for variability across trials. Fitting was done using the *fitlme* function in MATLAB (MathWorks, MA).

For Experiment 1, we modeled normalized RTs for each trial using three variables and all permutations of interactions between them as fixed effects, including target presence, object orientation, and set size. Random effects on intercept included participant identity and object category in each trial. The object orientation factor in the model was considered a measure of orientation dependency. The interaction between object orientation and set size was considered the effect of orientation on the slope of the RT-set size function. The three-way interaction between object orientation, set size, and target presence was considered the effect of orientation on the slope of the RT- effective set size function.

For Experiment 2, we modeled normalized RTs using object orientation and the orientations of visual context and egocentric reference frame as fixed effects, and the random effects were the same as in Experiment 1. Interactions between Object orientation x Visual context and Object orientation x Egocentric were considered

contributions of visual context and egocentric cues, respectively. The three-way interaction between Object orientation x Visual context x Egocentric was considered the effect of gravitational cues.

Contributions of reference frames. To quantify the relative weights of each reference frame in Experiment 2, we first computed an orientation dependency index (ODI) as the difference in z-scored RT for horizontal and upright objects:

$$ODI = z(RT_{90^\circ}) - z(RT_{0^\circ}).$$

Weights for reference frames were then computed as the change in ODI in each condition compared to the baseline:

$$w_{egocentric} = ODI_{baseline} - ODI_{egocentric}$$

$$w_{visual\ context} = ODI_{baseline} - ODI_{visual\ context}$$

$$w_{gravitational} = ODI_{baseline} - ODI_{gravitational}$$

Finally, the weights were normalized by their sum for each participant. Note that object orientation was defined relative to the reference frame that was not under manipulation in each experimental condition (i.e., gravitational reference frame in Baseline, Visual context, and Egocentric conditions; egocentric reference frame in Gravitational condition). A linear mixed-effects model was fit to the weights with participant identity as a random effect: Weight ~ Visual context + Egocentric + Gravitational + (1|Subject ID). This provides a statistical assessment of the significance of each weight. An analysis of variance on the model was used to measure variability across reference cues.

Reference

1 Treisman, A. & Souther, J. Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General* **114**, 285 (1985).

2 Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cognitive psychology* **12**, 97-136 (1980).

3 Egeth, H. & Dagenbach, D. Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance* **17**, 551 (1991).

4 Eckstein, M. P. Visual search: A retrospective. *Journal of vision* **11**, 14-14 (2011).

5 Wolfe, J. M. Approaches to visual search: Feature integration theory and guided search. *The Oxford handbook of attention*, 11-55 (2014).

6 Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I. & Sherman, A. M. Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics* **73**, 1650-1671 (2011).

7 Zhaoping, L. & Frith, U. A clash of bottom-up and top-down processes in visual search: The reversed letter effect revisited. *Journal of Experimental Psychology: Human Perception and Performance* **37**, 997-1006 (2011). https://doi.org/10.1037/a0023099

8 Li, Z. Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences* **96**, 10530-10535 (1999).

9 Wolfe, J. M. Guided Search 6.0: An updated model of visual search. *Psychonomic bulletin & review* **28**, 1060-1092 (2021).

10 Wolfe, J. M. & Horowitz, T. S. Five factors that guide attention in visual search. *Nature human behaviour* **1**, 0058 (2017).

11 Josephs, E. L., Draschkow, D., Wolfe, J. M. & Võ, M. L.-H. Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica* **169**, 100-108 (2016).

12 Beitner, J., Helbing, J., Draschkow, D. & Vo, M. L.-H. Get your guidance going: Investigating the activation of spatial priors for efficient search in virtual reality. *Brain Sciences* **11**, 44 (2021).

13 Lauer, T. & Võ, M. L.-H. The ingredients of scenes that affect object search and perception. *Human perception of visual information: Psychological and computational perspectives*, 1-32 (2022).

14 Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience* **14**, 926-932 (2011).

15 Howard, I. P. *Human visual orientation*. (John Wiley & Sons, 1982).

16 Appelle, S. Perception and discrimination as a function of stimulus orientation: the" oblique effect" in man and animals. *Psychological bulletin* **78**, 266 (1972).

17 Bülthoff, H. H., Edelman, S. Y. & Tarr, M. J. How are three-dimensional objects represented in the brain? *Cerebral Cortex* **5**, 247-260 (1995).

18 Jolicoeur, P. & Humphrey, G. K. Perception of rotated two-dimensional and three-dimensional objects and visual shapes. *Perceptual constancy. Why things look as they do*, 69-123 (1998).

19 Lawson, R. Achieving visual object constancy across plane rotation and depth rotation. *Acta psychologica* **102**, 221-245 (1999).

20 Tarr, M. J. Visual object recognition: Can a single mechanism suffice? (2003).

21 Tarr, M. J. & Bülthoff, H. H. Image-based object recognition in man, monkey and machine. *Cognition* **67**, 1-20 (1998).

22 Boutsen, L. & Marendaz, C. Detection of shape orientation depends on salient axes of symmetry and elongation: Evidence from visual search. *Perception & Psychophysics* **63**, 404-422 (2001).

23 Jolicoeur, P. Orientation congruency effects in visual search. *Canadian Journal of Psychology/Revue canadienne de psychologie* **46**, 280 (1992).

24 Xu, Z.-X., Chen, Y. & Kuai, S.-G. The human visual system estimates angle features in an internal reference frame: A computational and psychophysical study. *Journal of vision* **18**, 10-10 (2018).

25 Alexander, A. S., Robinson, J. C., Stern, C. E. & Hasselmo, M. E. Gated transformations from egocentric to allocentric reference frames involving retrosplenial cortex, entorhinal cortex, and hippocampus. *Hippocampus* **33**, 465-487 (2023).

26 Sasaki, R., Anzai, A., Angelaki, D. E. & DeAngelis, G. C. Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nature neuroscience* **23**, 1004-1015 (2020).

27 Palmer, S. E. in *Object perception* 121-163 (Psychology Press, 2013).

28 Chang, D. H., Harris, L. R. & Troje, N. F. Frames of reference for biological motion and face perception. *Journal of Vision* **10**, 22-22 (2010).

29 Galati, G., Pelle, G., Berthoz, A. & Committeri, G. Multiple reference frames used by the human brain for spatial perception and memory. *Experimental brain research* **206**, 109-120 (2010).

30 Troje, N. F. Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception* **32**, 201-210 (2003).

31 Andersen, R. A., Shenoy, K. V., Snyder, L. H., Bradley, D. C. & Crowell, J. A. The contributions of vestibular signals to the representations of space in the posterior parietal cortex. *Ann N Y Acad Sci* **871**, 282-292 (1999). https://doi.org/10.1111/j.1749-

6632.1999.tb09192.x

32 Snyder, L. H., Grieve, K. L., Brotchie, P. & Andersen, R. A. Separate body- and world-referenced representations of visual space in parietal cortex. *Nature* **394**, 887-891 (1998). https://doi.org/10.1038/29777

33 Marendaz, C. Nature and dynamics of reference frames in visual search for orientation: Implications for early visual processing. *Psychological Science* **9**, 27-32 (1998).

34 Brotchie, P. R., Andersen, R. A., Snyder, L. H. & Goodman, S. J. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **375**, 232-235 (1995). https://doi.org/10.1038/375232a0

35 Rosenberg, A. & Angelaki, D. E. Gravity influences the visual representation of object tilt in parietal cortex. *Journal of Neuroscience* **34**, 14170-14180 (2014).

36 Harris, L. R. *et al.* How our body influences our perception of the world. *Frontiers in psychology* **6**, 819 (2015).

37 Dyde, R. T., Jenkin, M. R., Jenkin, H. L., Zacher, J. E. & Harris, L. R. The effect of altered gravity states on the perception of orientation. *Experimental brain research* **194**, 647-660 (2009).

38 Dyde, R. T., Jenkin, M. R. & Harris, L. R. The subjective visual vertical and the perceptual upright. *Experimental Brain Research* **173**, 612-622 (2006).

39 Jenkin, H. L., Jenkin, M. R., Dyde, R. T. & Harris, L. R. Shape-from-shading depends on visual, gravitational, and body-orientation cues. *Perception* **33**, 1453-1461 (2004).

40 Lauer, T., Willenbockel, V., Maffongelli, L. & Võ, M. L.-H. The influence of scene and object orientation on the scene consistency effect. *Behavioural Brain Research* **394**, 112812 (2020).

41 Harris, L. R., Herpers, R., Hofhammer, T. & Jenkin, M. How much gravity is needed to establish the perceptual upright? *PLoS One* **9**, e106207 (2014).

42 Chen, X. & Zelinsky, G. J. Real-world visual search is dominated by top-down guidance. *Vision research* **46**, 4118-4133 (2006).

43 Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex (New York, NY: 1991)* **7**, 181-192 (1997).

44 Tusa, R., Palmer, L. & Rosenquist, A. The retinotopic organization of area 17 (striate cortex) in the cat. *Journal of Comparative Neurology* **177**, 213-235 (1978).

45 Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* **160**, 106 (1962).

46 Andersen, R. A. Visual and eye movement functions of the posterior parietal cortex. *Annu Rev Neurosci* **12**, 377-403 (1989).

https://doi.org/10.1146/annurev.ne.12.030189.002113

47 Andersen, R. A., Essick, G. K. & Siegel, R. M. Encoding of spatial location by posterior parietal neurons. *Science* **230**, 456-458 (1985). https://doi.org/10.1126/science.4048942

48 Peltier, N. E., Anzai, A., Moreno-Bote, R. & DeAngelis, G. C. A neural mechanism for optic flow parsing in macaque visual cortex. *Current Biology* **34**, 4983-4997. e4989 (2024).

49 Xu, Z.-X. & DeAngelis, G. C. Neural mechanism for coding depth from motion parallax in area MT: gain modulation or tuning shifts? *Journal of Neuroscience* **42**, 1235-1253 (2022).

50 Parker, P. R., Brown, M. A., Smear, M. C. & Niell, C. M. Movement-related signals in sensory areas: roles in natural behavior. *Trends in neurosciences* **43**, 581-595 (2020).

51 Warren, P. A. & Rushton, S. K. Optic flow processing for the assessment of object movement during ego movement. *Current Biology* **19**, 1555-1560 (2009).

52 Niehorster, D. C. & Li, L. Accuracy and tuning of flow parsing for visual perception of object motion during self-motion. *i-Perception* **8**, 2041669517708206 (2017).

53 Inaba, N., Shinomoto, S., Yamane, S., Takemura, A. & Kawano, K. MST neurons code for visual motion in space independent of pursuit eye movements. *J Neurophysiol* **97**, 3473-3483 (2007). https://doi.org/10.1152/jn.01054.2006

54 Xu, Z.-X., Pang, J., Anzai, A. & DeAngelis, G. C. Viewing geometry drives flexible perception of object motion and depth. *bioRxiv*, 2024.2010. 2029.620928 (2024).

55 Morris, A. P. & Krekelberg, B. A stable visual world in primate primary visual cortex. *Current Biology* **29**, 1471-1480. e1476 (2019).

56 Parker, P. R., Abe, E. T., Leonard, E. S., Martins, D. M. & Niell, C. M. Joint coding of visual input and eye/head position in V1 of freely moving mice. *Neuron* **110**, 3897-3906. e3895 (2022).

57 Miura, S. K. & Scanziani, M. Distinguishing externally from saccade-induced motion in visual cortex. *Nature* **610**, 135-142 (2022).

58 Marendaz, C., Stivalet, P., Barraclough, L. & Walkowiac, P. Effect of gravitational cues on visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance* **19**, 1266 (1993).

59 Whelan, R. Effective analysis of reaction time data. *The psychological record* **58**, 475-482 (2008).