

The reference frame for encoding and retention of motion depends on stimulus set size

Duong Huynh¹ · Srimant P. Tripathy² · Harold E. Bedell^{3,4} · Haluk Ögmen^{1,3,5}

Published online: 13 January 2017
© The Psychonomic Society, Inc. 2017

Abstract The goal of this study was to investigate the reference frames used in perceptual encoding and storage of visual motion information. In our experiments, observers viewed multiple moving objects and reported the direction of motion of a randomly selected item. Using a vector-decomposition technique, we computed performance during smooth pursuit with respect to a spatiotopic (nonretinotopic) and to a retinotopic component and compared them with performance during fixation, which served as the baseline. For the stimulus encoding stage, which precedes memory, we found that the reference frame depends on the stimulus set size. For a single moving target, the spatiotopic reference frame had the most significant contribution with some additional contribution from the retinotopic reference frame. When the number of items increased (Set Sizes 3 to 7), the spatiotopic reference frame was able to account for the performance. Finally, when the number of items became larger than 7, the distinction between reference frames vanished. We interpret this finding

as a switch to a more abstract nonmetric encoding of motion direction. We found that the retinotopic reference frame was not used in memory. Taken together with other studies, our results suggest that, whereas a retinotopic reference frame may be employed for controlling eye movements, perception and memory use primarily nonretinotopic reference frames. Furthermore, the use of nonretinotopic reference frames appears to be capacity limited. In the case of complex stimuli, the visual system may use perceptual grouping in order to simplify the complexity of stimuli or resort to a nonmetric abstract coding of motion information.

Keywords Motion perception · Reference frame · Sensory memory · Iconic memory · Short-term memory

Electronic supplementary material The online version of this article (doi:10.3758/s13414-016-1258-5) contains supplementary material, which is available to authorized users.

✉ Haluk Ögmen
ogmen@du.edu

¹ Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA

² School of Optometry and Vision Science, University of Bradford, Bradford, UK

³ Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA

⁴ College of Optometry, University of Houston, Houston, TX, USA

⁵ Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80208, USA

The optics of the eyes map the three-dimensional visual scene into two-dimensional retinal images. The projections from retina to subcortical and to early visual cortical areas preserve the neighborhood relationships in these images, creating what is known as *retinotopic representations* (Engel, 1994; Gardner, Merriam, Movshon, & Heeger, 2008; Sereno, Pitzalis, & Martinez, 2001; Tootell et al., 1998; Tootell et al., 1995). Retinotopic representations are informative about the position of stimuli in the scene with respect to the eyes and hence can play a role in the control of eye movements (McKenzie & Lisberger, 1986). For example, to make a saccade to a selected stimulus, sensorimotor systems can compute the “error signal”—that is, the distance of the target with respect to fovea, within retinotopic representations and use the error signal to program the movements of the eyes (McKenzie & Lisberger, 1986; Orban de Xivry & Lefèvre, 2007). There is also evidence that spatiotopic representations (i.e., representations based on reference frames that are located in space) contribute to the control of eye movements (Mays & Sparks,

1980; Pertzov, Avidan, & Zohary, 2011). On the other hand, retinotopic representations are not well suited to explain our perceptual experience under normal viewing conditions due to the movements of the observer (eye, head, body movements, which we call “ego-motion”) and those of the objects in the environment (which we call “exo-motion”).

Natural human vision is based on a sequence of saccadic or smooth gaze changes that direct the fovea to and maintain it on stimuli of interest (Buswell, 1935; Yarbus, 1967; Zelinsky & Todor, 2010). These ego-motions cause drastic shifts of stimuli in retinotopic representations. Despite this instability in retinotopic representations, our perceptual experience of the visual world appears to be highly stable. It stands to reason that other coordinate systems, that we will call collectively *nonretinotopic representations*, are necessarily involved for the visual system to achieve a sense of spatiotemporal coherence (Bridgeman, Van der Heijden, & Velichkovsky, 1994; Burr & Morrone, 2011, 2012; Cavanagh, Hunt, Afraz, & Rolfs, 2010; Melcher & Colby, 2008; Melcher & Morrone, 2015; Wurtz, 2008). It has been suggested that efference-copy signals associated with ego-motion commands play an important role in transforming retinotopic representations into nonretinotopic representations (Andersen, Snyder, Li, & Stricanne, 1993; Bridgeman, 1995; Mack, 1986; Von Helmholtz, 1925; Von Holst, 1954). In addition to stabilizing our percepts, the inclusion of efference-copy signals to build up nonretinotopic representations may also improve abilities such as localization of speed differences in complex stimuli. For example, Braun, Schütz, and Gegenfurtner (2010) measured thresholds for detecting the spatial location of stimuli that undergo speed changes during fixation and smooth pursuit. They showed that the ability to spatially localize speed changes was better during pursuit compared to fixation, which they attributed to the use of efference-copy signals (Braun et al., 2010).

Another problem for retinotopic representations stems from the movements of objects in the environment (Burr, 1980; Chen, Bedell, & Ögmen, 1995; Nishida, 2004; Ögmen, 2007; Ögmen & Herzog, 2010). When objects in the environment move (exo-motion), they activate retinotopically anchored mechanisms only briefly and hence the resulting percepts are predicted to be blurred with “ghost-like” appearances (Ögmen, 2007; Ögmen & Herzog, 2010). However, our percepts of moving objects are in general sharp and clear. In the case of exo-motion, it has been suggested that the motion of objects are used to build reference frames to transform retinotopic representations into nonretinotopic representations (Bremner, Bryant, & Mareschal, 2005; Duncker, 1929/1938; Ögmen, Herzog, & Noory, 2013; Wade & Swanston, 1987, 1996).

In addition to the question of which reference frames are used to explain our motor behavior and our real-time perceptual experience, it is also important to understand the reference frames used in memory systems. According to the modal

model of human memory (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974), the visual stimulus-encoding stage is followed by three memory systems: visual-sensory memory (VSM), visual-short-term memory (VSTM), and long-term memory (LTM).

Whereas much of the research on VSM indicates that it is encoded in retinotopic coordinates (e.g., Haber, 1983; Jonides et al., 1983; Rayner & Pollatsek, 1983; Irwin et al., 1983, 1988; Sun & Irwin, 1987), more recent findings using sequential metacontrast and Ternus-Pikler displays indicate that sensory memory can also employ motion-based nonretinotopic reference frames (Ögmen, 2007; Ögmen & Herzog, 2010; Noory, Herzog, & Ögmen, 2015).

Studies on the reference frames underlying VSTM have been equivocal. In a study by Baker, Harper, and Snyder (2003), monkeys were trained to hold in memory either the retinotopic or spatiotopic location of a stimulus, and then, after a slow gaze shift (smooth pursuit eye movement), make a saccade toward the remembered location. They found that saccades to spatiotopic locations are more variable than saccades to retinotopic locations, and suggested a retinotopically organized model of VSTM for spatial information. Retinotopic processing in VSTM was also suggested by Golomb and Kanwisher (2012) based on the finding that, after making a visually guided saccade during a memory-delay interval, observers were significantly more accurate and precise at reporting retinotopic locations than spatiotopic locations. Results from study by Ong, Hooshvar, Zhang, and Bisley (2009), on the contrary, favor the spatiotopic model of VSTM. In a delayed-saccade paradigm, observers were asked to compare directions of motion of a presaccadic and a postsaccadic stimulus. Performance was found optimal when the two stimuli appear at the same spatiotopic, rather than retinotopic, location.

Given the important role that motion information carries in visual processing and given that motion itself can constitute a reference frame, the goal of our study was to investigate the reference frames used in the encoding and the retention of motion information in VSM and VSTM. In principle, to distinguish retinotopic and nonretinotopic components of visual processing, one has to manipulate movements of the observer (e.g., the eyes) and of the objects, because the two systems are not separable under static viewing. Most of the previous work investigating reference frames for visual memory reviewed earlier was based on procedures that involved saccadic eye movements. The use of only vertical and/or horizontal saccadic displacements of the eyes and highly predictable locations of objects might limit the generality of those findings. In addition, although visual stimuli were defined on the basis of attributes such as orientation or motion, their spatial location was the only factor used to contrast retinotopic and nonretinotopic conditions. In this study, with eye displacements incorporated in the form of smooth-pursuit eye

movements (SPEMs), we dissociated the two reference frames by applying motion-vector decomposition (see Data Analysis section). Directions of eye and object movements were both randomized in our experiments. We sought to investigate the coordinates in which the visual system encodes and stores in memory the directions of motion of multiple moving stimuli.

Previous studies addressing reference frames for motion stimuli mostly focused on processing at the perceptual encoding level. The perceived direction of motion during pursuit has been reported to have retinotopic (Becklen, Wallach, & Nitzberg, 1984; Festinger, Sedgwick, & Holtzman, 1976; Mateeff, 1980; Wallach, Becklen, & Nitzberg, 1985) and incompletely converted spatiotopic coordinates (Souman et al., 2005a, 2005b; Souman et al., 2006a, 2006b; Swanston & Wade, 1988). In terms of attentively tracking the identities of multiple moving objects, it has been suggested that both retinotopic and spatiotopic coordinate systems are used (Howe, Pinto, & Horowitz, 2010). A more recent study showed that the effective reference frame for motion consists of an integration of motion-based, retinotopic and spatiotopic reference frames (Agaoglu, Herzog, & Ögmen, 2015a, b). However, very few studies have investigated reference frames underlying memory for motion. Melcher and Fracasso (2012) investigated the trans-saccadic line-motion illusion (TLMI) by introducing a saccade between the presentation of the inducer and the line. They also presented two inducers on each trial such that any illusion operating in a retinotopic reference frame would be in opposition to any illusion in a spatiotopic reference frame. The study found that the direction of the TLMI perceived was largely consistent with a spatiotopic reference frame. In a separate experiment the authors varied the number of inducers in the TLMI stimulus and found that observers had a capacity of approximately two inducers, well below the capacity of about seven that they found in a comparable transsaccadic visual working memory experiment for color. They suggested the trans-saccadic capacity was limited by the number of object files or attentional pointers that could be updated across saccades. It is not clear how this TLMI study would generalize to other motion tasks, because in TLMI the inducers used in the memory task were stationary, as was the line, and the percept in line motion illusion differs from that in apparent motion, though there is some overlap of the neural mechanisms involved in the two tasks (e.g., Jancke, Chavane, Naaman, & Grinwald, 2004).

In this study, using a partial-report technique, in which the cue was delivered immediately, or else with varying delays, after stimulus offset, we examined perception of motion in the different processing stages, from encoding to sensory memory and VSTM. We determined reference systems for each stage of motion processing in two conditions, with and without eye movements. With eye movements (SPEM condition), nonretinotopic and retinotopic coordinates are dissociable. In general, if motion is processed primarily in one coordinate

system, performance measured in that system during SPEM should be better than that measured in the other system. Also, the former should be comparable to the performance level in the absence of eye movements (fixation condition, in which case, nonretinotopic performance is the same as retinotopic performance).

Previous studies have varied set size to get estimates of capacity of the visual system for processing multiple motions in a variety of tasks involving: tracking of object identities (Pylyshyn & Storm, 1988); monitoring changes in direction of motion (Tripathy & Barrett, 2004; Tripathy, Narasimhan, & Barrett, 2007; Narasimhan, Tripathy, & Barrett, 2009); monitoring directions of motion (Horowitz & Cohen, 2010; Shooner, Tripathy, Bedell, & Ögmen, 2010); encoding and memory of directions of motion (Ögmen, Ekiz, Huynh, Tripathy, & Bedell, 2013); and feature binding (Huynh, Tripathy, Bedell, & Ögmen, 2015). An open question is whether a single reference frame is used for processing motion across all set sizes. Several recent studies point to the existence of multiple complementary mechanisms for processing multiple objects, in particular, low-capacity systems that can process three or four items individually, and high-capacity systems for encoding and storing summary statistics of sets/groups/ensembles of objects larger than the aforementioned three or four items (Cant, Sun, & Xu, 2015). The summary statistics that are extracted during ensemble coding include means or averages (reviewed in Bauer, 2015) and variances (Norman, Heywood, & Kentridge, 2015). Ensemble coding has been demonstrated for a variety of features including low-level ones such as mean size (Ariely, 2001; Corbett & Melcher, 2014), variance in orientation (Norman et al., 2015), and high-level features such as averages of emotions, gender and identities of faces, and behaviors of crowds (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007; Sweeny, Haroz, & Whitney, 2012). It is likely that different mechanisms are involved in the extraction of summary statistics of different features (Hubert-Wallander & Boynton, 2015), making generalizations difficult. Studies of ensemble encoding usually require observers to report some average statistic of some stimulus feature, rather than some feature pertaining to one particular object as in our partial report experiments. However, some contribution of ensemble encoding to the measured capacity cannot be ruled out (Brady & Alvarez, 2015). We varied set size between one and 12 in order to estimate the capacity of the motion system for representing direction of motion at the encoding and memory stages. The range of set sizes permits us to investigate if different mechanisms operate at small and large set sizes, using potentially different reference frames.

Statistical modeling further allowed us to determine whether a purely nonretinotopic, retinotopic, or a combined model best describes the behavioral data, as well as to probe the quantitative and qualitative details of observers' performance.

Method

We ran a set of four experiments in which observers tracked multiple moving objects while (1) maintaining their gaze on a fixation point (Experiments 1a–1b); or (2) performing a smooth pursuit eye movement (SPEM; Experiments 2a–2b). The task of the observers was to report the perceived direction of motion of a randomly chosen object by rotating an on-screen pointer. Experiments 1a and 2a, in which the target to be reported was cued immediately after objects stopped moving and disappeared, aimed to characterize the initial encoding stage. While observers had to hold in memory information about this cued target during the reporting phase, having a single target item and no delay after stimulus offset minimized the involvement of memorization in their performance. Experiments 1b and 2b included varying cue delays, aimed to tap into sensory memory and VSTM. For each eye movement condition, we provided observers with several initial training blocks (28 trials each) to ensure that each of them could perform all tasks well. In general, a reasonable proportion of valid trials (>70%) was obtained after two or three such blocks. Criteria for the validity of a trial in the fixation and SPEM conditions are described under the Procedure sections of Experiments 1a and 2a, respectively. These criteria were applied for only the dominant eye of each observer, which was determined in advance using the ABC test for sighting dominance (Miles, 1929, 1930). We used the dominant eye because subjects fixate significantly more accurately with the dominant eye compared to the nondominant eye during eye-position calibration (Nyström, Andersson, Holmqvist, & van de Weijer, 2013; Vikesdal & Langaag, 2016) and track more faithfully target motion with the dominant eye (Gibaldi, Canessa, & Sabatini, 2016). One participant had a left dominant eye (TTN), the remaining three had right dominant eyes (TAN, QVP, DHL). Eye positions for both eyes were recorded, but only the dominant eye's data were used in analyses. In addition, a short training block (seven or 14 trials) was also run when observers came back after a break to foster stability of performance.

Participants

The first author and three naïve observers with normal, or corrected to normal visual acuity, and with no color deficiency (according to self-reports and the online version of the Ishihara test) participated in all experiments. Naïve observers were not informed about the hypotheses of the study. Experiments were conducted according to a protocol adhering to the Declaration of Helsinki and approved by the University of Houston Committee for the Protection of Human Subjects.

Apparatus

A Visual Stimulus Generator system (Cambridge Research Systems) with a VSG2/5 video card housed in a personal

computer and a SONY GDM-FW900 color monitor (20 inches, 100 Hz) were used to create and display stimuli; programming was implemented in C++. The screen resolution was 800×500 pixels of which 604×405 pixels (19.7×13.2 deg; 1.96 arcmin/pixel) were used for object display. The screen edges were visible during the experiments, but the border of the display area was not. Observers used a computer mouse to give their response, and their heads were kept still on a head/chin rest at a distance of 1 meter in front of the monitor. Gaze position and velocity were recorded using a head-mounted binocular eye-tracking system (SR Research, Eyelink II) sampling at 250 Hz.

Stimuli

A black cross subtending 1.5×1.5 deg was used to guide eye fixation or pursuit movements. Objects were circular disks of different readily distinguishable colors that were randomly selected from a set of 180 equi-luminant colors. These 180 colors were sampled along a circle (i.e., resolution $2^\circ/\text{color}$) in the CIE L^*a^*b color system. The circle is located at $L = 15 \text{ cd/m}^2$, centered at the white point (with $a = 0.2044$ and $b = 0.4808$), and its radius was chosen to maximize the discriminability of the colors (approximately 2°). Color separation of any two objects was not smaller than 17° (see Huynh et al., 2015, for a justification of this separation). Although color was task irrelevant in the present experiments, we used colored instead of uniform gray objects, as in the Huynh et al. study, for comparison purposes. The diameter of each object was chosen to subtend a visual angle of 1 deg. Objects were presented on a gray 40-cd/m^2 background.

Experiment 1a: Reference frames for stimulus encoding during fixation

Procedure A trial began with a fixation-cross presented at the center of the screen. Observers were instructed to start fixating the cross when ready and promptly click the mouse. Upon detecting this mouse click, the program sent a trigger code to the eye tracker. An online drift correction was performed and eye movement recordings started at this point. Observers continued to hold fixation on the cross. At 1,300 ms after the mouse click,¹ a stimulus containing multiple moving disks was displayed for 200 ms while the fixation cross remained stationary. The disks moved along linear trajectories in random directions at a speed of 5 deg/s. To minimize interference between the objects, the disk trajectories were constrained never to cross one another and no two objects had motion directions closer than 17 deg. After 200 ms of presentation, the stimuli were removed from the display. One of the disks was randomly chosen to be the probed item, the position of

¹ The choice of this duration was to make it consistent with the SPEM condition to be described later.

which was cued by a small black dot, and observers were asked to report the disk's motion direction by rotating an on-screen pointer, which was a black bar extending from the dot, to the perceived direction (see Fig. 1). Although the fixation cross remained visible until response, eye tracking ended and fixation was no longer strictly required after the offset of the stimuli.

Note that the present experiments were different than our previous experiments in Huynh et al. (2015) in that (1) No static preview of objects (1 s) was shown before the motion period as we wanted the objects to only appear during the steady phase of smooth pursuit, and (2) no feedback was provided after responses in either the training or experimental runs. Because gaze position was controlled in the present experiments, observers might learn from feedback and adjust their responses to avoid errors that they realized they made repeatedly when the target moved across their gaze point at certain distances or directions.

As mentioned earlier, gaze position was monitored by the eye tracker. During each trial, observers were asked to maintain their gaze on the fixation point at the center of the screen until the disk stimuli disappeared. To be considered as a good fixation, observers' gaze had to remain within a circular area with 1-deg radius around the central point of the cross, with no saccades or blinks. If any of these requirements were not met, the trial was rejected and observers received text feedback on the display telling the reason for rejection. The rejected trial was repeated later during the same set of trials.

Design Seven set sizes (one, three, four, six, eight, nine, or 12) of moving objects were tested. The experiment was divided into 25 separate blocks with trials of all set sizes randomly

interleaved within each block. A block ended whenever observers finished 28 valid trials (four trials per condition of set size). That is, each observer ran $4 * 25 = 100$ trials per set size, or 700 trials in total. The eye tracker was recalibrated at the start of each block using a nine-point grid. Calibration validation was checked twice, after the calibration and when the observer completed a block. If any of the fixated positions during validation disagreed from the original calibration by more than 1.5 deg at the conclusion of any block, the entire block was excluded and rerun later.

Experiment 1b: Reference frames for sensory memory and VSTM during fixation

Experiment 1b was the same as Experiment 1a except for the following changes:

- The number of objects was fixed at six in every block.
- The cue was not always given immediately after the objects disappeared, but was preceded by a variable-duration delay. Seven different delay values (0, 50, 100, 250, 500, 1000, or 3000 ms) were randomly chosen on each trial.

As in Experiment 1a, observers finished a block after obtaining 28 valid trials (four trials per condition of cue delay). A total of 25 blocks yielded $4 * 25 = 100$ trials per condition of cue delay, or 700 trials in total. Observers followed the same steps as in Experiment 1a. Invalid trials and blocks were discarded and rerun.

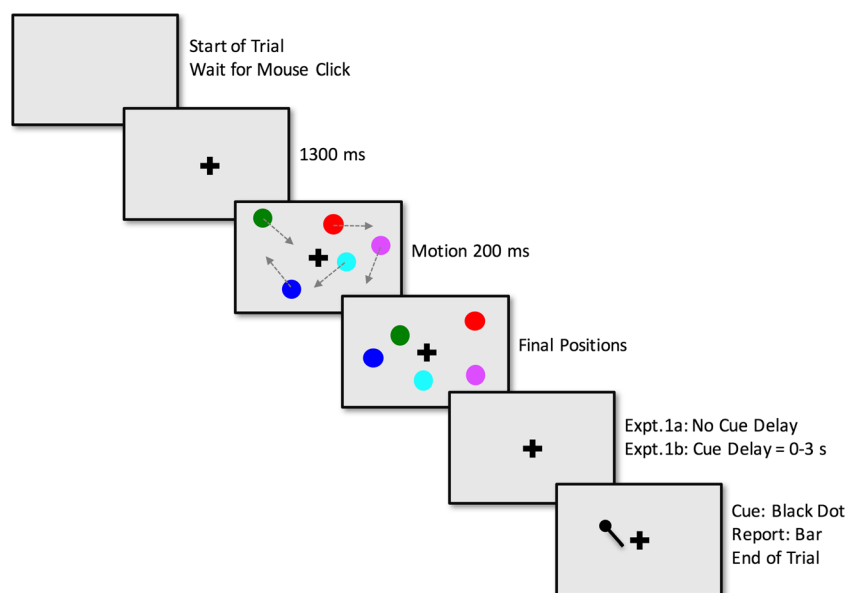


Fig. 1 Time course of a trial in the fixation condition: Experiments 1a (varying set size; no cue delay) and 1b (set size fixed; varying cue delay) (Color figure online)

Experiment 2a: Reference frames for stimulus encoding during SPEM

Procedure We applied the step-ramp paradigm devised by Rashbass (1961) to obtain a relatively fast and smooth initiation of pursuit eye movement (see Fig. 2). A cross of the same design as in the fixation condition served as the pursuit target. Observers were required to fixate initially and then smoothly follow this target as it moved. At the beginning of each trial, the cross was presented at a randomly selected location on an invisible circle that was centered at the center of the screen and had a radius of 3 deg (in terms of visual angle). Observers were instructed to start fixating the cross when ready and promptly click the mouse. A command was then sent by the program to request the eye tracker to perform drift correction and start recording eye data. The cross remained stationary for 500 ms after the mouse click, then suddenly jumped in the centrifugal direction to another location which was at 4 deg away from the center of the screen (step size = 1 deg), and immediately started moving in the opposite direction toward the center at a constant speed (5 deg/s). The target reached the center of the display at 800 ms after the step and continued to move in the same direction for an additional 200 ms. The stimuli containing multiple moving disks were displayed during this 200-ms period (Fig. 3). Similar to the fixation condition, the disks moved along linear trajectories in random directions at the speed of 5 deg/s, the same as that of the pursuit target. When the pursuit target stopped moving at 1 deg from the center point, the stimuli were removed from the display. A

randomly chosen disk was marked immediately with a small black dot to indicate the target for response. The task of reporting the target’s direction of motion was the same as in Experiments 1a and 1b.

Criteria for a valid smooth-pursuit trial To guarantee that observers successfully performed smooth-pursuit eye movements while tracking the moving disks, a number of requirements had to be met. First, during the initial fixation phase, observers had to maintain their gaze within an invisible circle (1.5 deg radius) around the center of the cross. Second, given the pursuit latency on the order of 120 ms, pursuit onset had to be detected in the interval 120–600 ms after the step (see Fig. 2). Eye velocity (\vec{v}_e) had to exceed 25% of target velocity (\vec{v}_t) to be considered as pursuit onset. Third, we considered pursuit quality during the last 300 ms of pursuit, including the 200 ms of stimulus presentation and the previous 100 ms (Fig. 2). An EyeLink II built-in function was employed to calculate average eye velocity. This function allowed for the selection of the width of a moving window containing a certain group of most recent gaze-position samples that would be considered to estimate the velocity of the middle sample in the group, hence minimizing noise. In our pursuit experiments, we used the FIVE_SAMPLE_MODEL (width = 5 samples); with 4 ms/sample (EyeLink sample rate = 250 Hz). Therefore, five samples corresponded to 20 ms. By calling the function after each new sample, we obtained a running record of velocity for all samples during the last 300 ms of pursuit. Mean eye velocity over this

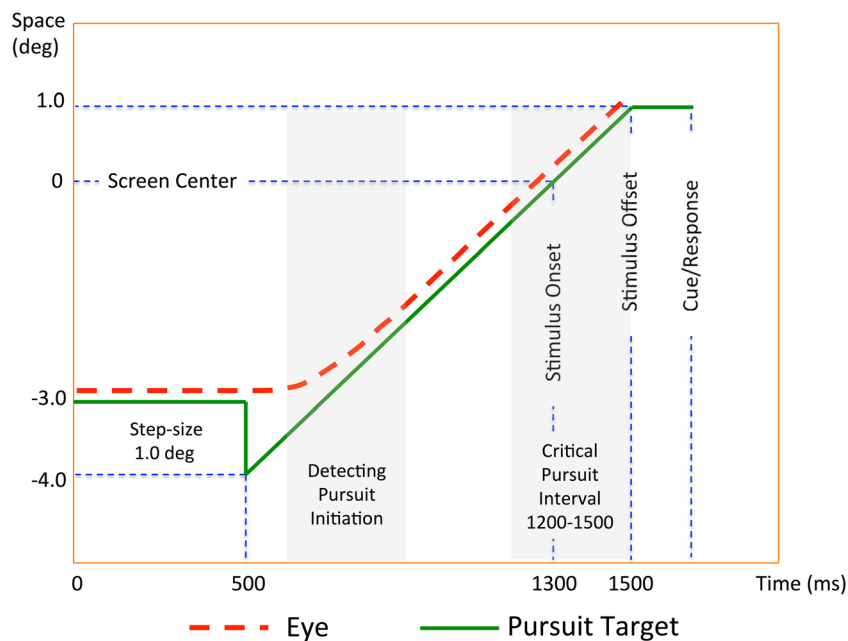


Fig. 2 Illustration of the step-ramp paradigm: changes of the pursuit target and eye position along the pursuit direction as a function of time (Color figure online)

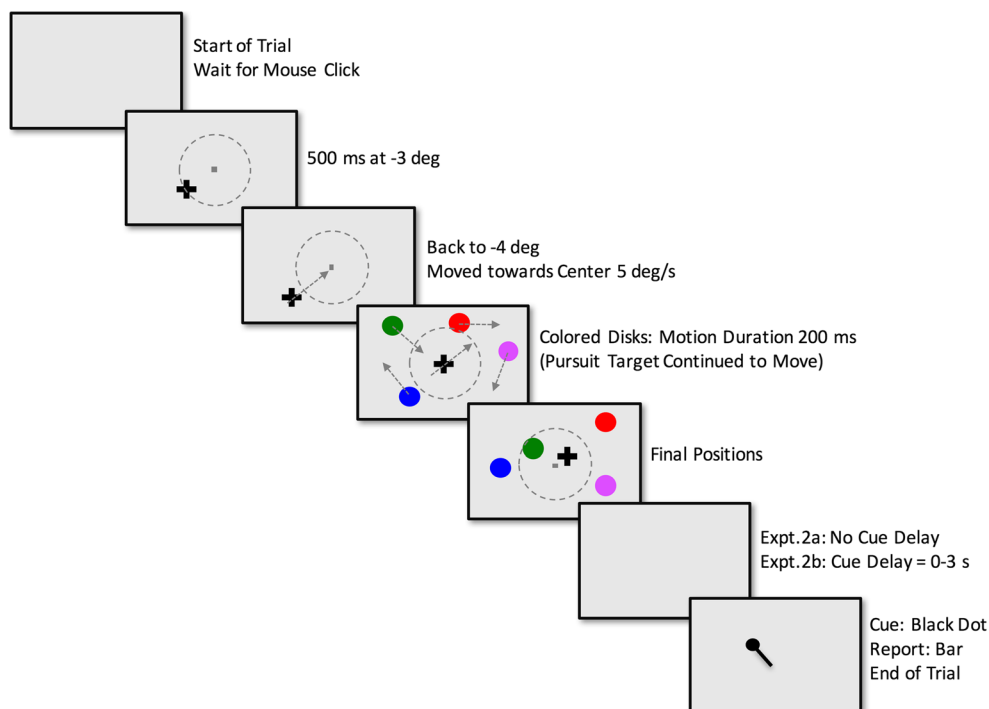


Fig. 3 Time course of a trial in the smooth pursuit eye movement (SPEM) condition: Experiments 2a (varying set size; no cue delay) and 2b (set size fixed; varying cue delay). The *gray central dot*, *dashed circle*,

and *dashed arrows* are shown for illustration purposes only; they are invisible during the experiments (Color figure online)

interval was then computed separately for the horizontal ($|\vec{v}_{ex}|$) and vertical ($|\vec{v}_{ey}|$) components. To qualify as a smooth pursuit, pursuit gain ($PG = |\vec{v}_e|/|\vec{v}_i|$) and that for either of the two components ($PG_x = |\vec{v}_{ex}|/|\vec{v}_{ix}|$ or $PG_y = |\vec{v}_{ey}|/|\vec{v}_{iy}|$) had to fall in the range [0.7, 1.3]. We did not put this constraint on both PG_x and PG_y due to the fact that, when the direction of \vec{v}_i was close to vertical or horizontal (rather small values of $|\vec{v}_{ix}|$ or $|\vec{v}_{iy}|$), it was virtually impossible to have the smaller component (PG_x or PG_y) fall in the specified range. Finally, also during the last 300 ms of pursuit, no saccades (saccade displacement threshold = 0.5 deg; saccade velocity threshold = 30 deg/s) or blinks were accepted. If any of the four constraints were violated, the trial was discarded and rerun during the same set of trials.

Design. Except for the difference in eye movement (smooth pursuit instead of fixation) and the corresponding criteria considered, the design of this experiment was the same as in Experiment 1a, with seven set sizes and 100 valid trials per condition of set size. Observers also received text feedback about eye movement after each rejected trial, which was rerun in the same block of trials. Similar to Experiments 1a and 1b, if postblock position calibrations disagreed with initial calibration values by more than 1.5 deg, the entire block was excluded and rerun later.

Experiment 2b: Reference frames for sensory memory and VSTM during SPEM

Experiment 2b was the same as Experiment 2a except for the following changes:

- The number of objects was fixed at six in every block.
- The cue was not given always immediately after the objects disappeared but was preceded by a variable-duration delay. Similar to Experiment 1b, seven different delay values (0, 50, 100, 250, 500, 1000, or 3000 ms) were randomly chosen on each trial.

Again, observers ran 25 blocks to obtain 100 valid trials per condition of cue delay.

Data analysis Our goal was to use statistical models to break the observers' aggregate performance down into multiple components that characterize important aspects of their behaviors. This includes consideration of the extent to which correct target reporting, guessing, and nontarget misreporting account for variability of response errors, and the nature of the reference frames associated with these errors. We wished to obtain both qualitative and quantitative measures for each of these components. We analyzed and compared several plausible models that are different from one another in their assumptions about an observer's

behavioral pattern. Our interpretations of the data were then based on the best performing model.

We used two different methods of fitting the models to empirical data (Huynh et al., 2015): (1) The method of least squares fitting involves the creation of a nonlinear optimization routine using the MATLAB *fminsearch*(.) function to find the values of the parameters that minimize an error function; (2) the expectation-maximization (E-M) algorithm (Dempster, Laird, & Rubin, 1977), which employs Bayes’ theorem for finding maximum likelihood estimates of the parameters. We found the best performing model by comparing adjusted R^2 (first method) and Akaike/Bayesian information criterion (second method) values obtained for each model and observed similar results. We report only the results of the first method in the main text. Mathematical derivation and results of the second method are provided in [Supplemental Information](#).

The first four of our hypothetical models (models F1, F2, F3c, F3r) are the same as in our previous study (Huynh et al., 2015). These models take into account noise and uncertainties in an observer’s responses and are independent of reference frames. They were used to analyze the fixation condition because stimulus motion is the same according to retinotopic and nonretinotopic coordinates in this condition. In the smooth pursuit condition, reference frame is incorporated into the models above as an additional factor. As shown in Fig. 4, the motion of the stimulus is analyzed in terms of a nonretinotopic and a retinotopic reference frame. The nonretinotopic reference frame refers to the motion of the stimulus on the monitor display and we refer to this as the *spatiotopic coordinates*. By taking into account the eye movement velocity, the motion of the stimulus can also be represented as a retinotopic motion-vector, which is referred to as *retinotopic coordinates*. We considered three different scenarios: (a) motion is processed only in spatiotopic coordinates (models SP1_S, SP2_S, SP3c_S, SP3r_S), (b) motion is processed only in retinotopic coordinates (models SP1_R,

SP2_R, SP3c_R, SP3r_R), and (c) both spatiotopic and retinotopic coordinates are used, or perhaps there is a gradual transition from one to the other such that they become active simultaneously (models SP1_SR, SP2_SR, SP3c_SR, SP3r_SR).

We present the models used in our analyses below; the detailed explanations of these models are provided in the [Supplemental Information](#) section. Briefly, the error distribution obtained from the data is fitted by an embedded family of statistical distributions. The member of the family that provides the “best fit” is selected to interpret the data. As explained below, parameters of the model are interpreted in terms of observer’s accuracy, precision, guess rate, and misbinding errors.

Fixation-condition models

Model F1: Gaussian This model is the cumulative distribution function of a circular (wrapped) Gaussian:

$$CDF(\varepsilon) = CDF\{G(\varepsilon; \mu, \sigma)\}, \tag{1}$$

where the cumulative distribution function $CDF(\varepsilon)$ of the error variable ε (ε = reported direction of motion - actual direction of motion) is given by a Gaussian distribution $G(\varepsilon; \mu, \sigma)$ whose parameters represent the accuracy (mean: μ) and the precision ($1/\sigma$, where σ is the standard deviation) of processing. The precision parameter $1/\sigma$ captures the qualitative aspect of performance, with smaller values of σ corresponding to higher qualities of encoding for the processed items.

Model F2: Gaussian + Uniform In this model (Zhang & Luck, 2008), the distribution of errors is represented by:

$$CDF(\varepsilon) = CDF\{w.G(\varepsilon; \mu, \sigma) + (1-w).U(-180, 180)\}, \tag{2}$$

where the cumulative distribution function $CDF(\varepsilon)$ is obtained

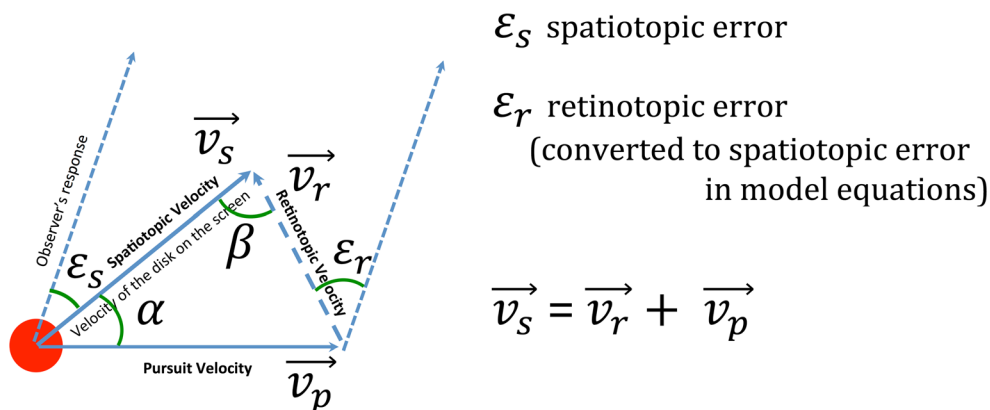


Fig. 4 Decomposition of spatiotopic and retinotopic motion vectors: \vec{v}_s , \vec{v}_r , and \vec{v}_p represent the velocity of the target object with respect to the screen (spatiotopic vector), velocity of the target object with respect to the projected fovea (retinotopic vector), and pursuit (eye) velocity,

respectively. Spatiotopic error is measured as the angular deviation between the reported and the spatiotopic vectors. Retinotopic error is measured as the angular deviation between the reported and the retinotopic vectors (Color figure online)

from the corresponding probability density function that consists of two components:

- (a) A Gaussian distribution $G(\varepsilon; \mu, \sigma)$ described in the Gaussian model
- (b) A uniform distribution U over the interval $(-180, 180)$, which represents guessing

The weight of the uniform distribution $(1 - w)$ represents the proportion of trials in which observers base their responses on guesses rather than on the target information available. The weight w of the Gaussian captures the quantitative aspect of performance by providing a relative measure for the *intake* of encoding, with a larger value corresponding to a greater possibility that a response is based on having some access to information from the cued target. Traditionally, the term *capacity* is used in the literature, as opposed to *intake*. By definition, capacity refers to the *maximum* amount of information that can be processed and/or stored. Hence, capacity refers in general to a *fixed* property of the system. Implicit in the definition of capacity is the idea that performance is unaffected by set size when it is smaller than the capacity. This condition does not hold for the perception of motion direction, or changes in the direction of motion, where substantial drop in precision for reporting direction of motion (or increase in threshold when detecting deviations) is seen with increases

of set size, even for set sizes of one or two (Huynh et al., 2015; Levi & Tripathy, 2006; Narasimhan et al., 2009; Ögmen et al., 2013; Tripathy & Barrett, 2004; Tripathy & Levi, 2008; Tripathy et al., 2007; Shooner et al., 2010). This drop in performance with set size necessitates an alternative way of characterizing the amount of information processed and/or stored in a given condition. For this purpose, we use the term *intake*, which represents the quantity of information processed/stored under a given stimulus condition (e.g., set size). As an analogy, the capacity of a room can be 50 people (i.e., the maximum number of people in the room), whereas under a given situation the room may be holding only 26 people (intake). As mentioned above, research indicates that performance decreases with set size in a continuous manner and a single capacity parameter is not adequate to characterize the amount of information that is processed and/or stored. Instead, using two parameters, one representing the variable *quantity* of information (intake) and a second one representing the *quality* of information (precision), appears to be a better theoretical approach (see, e.g., the “leaky flask” model in Ögmen et al., 2013).

Models F3c and F3r: Gaussian + Uniform + Gaussian

These models (Bays et al., 2009) include an additional term to account for misbinding errors when observers get confused and report another object instead of the selected target:

$$\text{CDF}(\varepsilon) = \text{CDF}\left\{w \cdot G(\varepsilon; \mu_t, \sigma_t) + (1-w-w_m) \cdot U(-180, 180) + w_m \cdot S_{i=1; i \neq t}^T [G(\varepsilon; \mu_t + \varepsilon_{i,t}, \sigma_t)]\right\}, \quad (3)$$

where the first two terms represent the same Gaussian and Uniform distributions as in the Gaussian + Uniform model and the third term represents errors stemming from misbinding reports. The selection operator $S_{i=1; i \neq t}^T[\cdot]$ determines which item from the set of $(T-1)$ noncued objects is the one that generates the subject’s response due to a misbinding error. We analyzed two versions of this model—that is, misbinding with the object that is closest to the cued target in either the cued-feature space (*closest cued feature*: Position – model F3c) or the reported-feature space (*closest reported feature*: Motion direction – model F3r).

Smooth-pursuit condition models

For all models in this section, the error variable is denoted by ε_s to emphasize the fact that we consistently computed errors in spatiotopic coordinates. On the circular ring that represents all possible values of motion direction, the actual motion direction of the target coincides with the origin of the spatiotopic coordinate system. A conversion parameter is included where necessary to convert spatiotopic errors to equivalent retinotopic

errors. The equations are similar if one prefers to compute errors in retinotopic coordinates, but the sign of the conversion parameter needs to be reversed.

Model SP1_S: Spatiotopic Gaussian This model has the same form as model F1:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{G(\varepsilon_s; \mu_s, \sigma_s)\}, \quad (4)$$

where μ_s and $1/\sigma_s$ respectively represent the accuracy and precision of spatiotopic processing.

Model SP1_R: Retinotopic Gaussian This model also consists of the CDF of a circular Gaussian:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{G(\varepsilon_s; \mu_r + \beta, \sigma_r)\}, \quad (5)$$

where μ_r and $1/\sigma_r$ respectively represent the accuracy and precision of retinotopic processing. The model assumes statistical analysis of retinotopic errors to produce a probability density function peaking near the actual retinotopic direction of motion and decaying for larger magnitudes of error.

However, because the error variable ε_s is calculated with respect to the actual spatiotopic direction of motion, the mean of the Gaussian must be shifted from the origin by an angle β determined by the difference between the actual spatiotopic and retinotopic directions. This angle is given by (see Fig. 4):

$$\beta = \sin^{-1} \left[\frac{|\vec{v}_p|}{|\vec{v}_r|} \sin \alpha \right], \tag{6}$$

where $|\vec{v}_p|$ and $|\vec{v}_r|$ are the magnitudes of the pursuit and retinotopic motion vectors, respectively, and α is the angle between spatiotopic and pursuit motion vectors.

Model SP1_SR: Spatiotopic Gaussian + Retinotopic Gaussian This model is the CDF of a weighted sum of two circular Gaussians:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{w_s \cdot G(\varepsilon_s; \mu_s, \sigma_s) + (1-w_s) \cdot G(\varepsilon_s; \mu_r + \beta, \sigma_r)\}, \tag{7}$$

where the weights w_s and $(1 - w_s)$ represent the relative contributions (or *intakes*) of spatiotopic processing (with accuracy μ_s and precision $1/\sigma_s$) and retinotopic processing (with accuracy μ_r and precision $1/\sigma_r$), respectively. The means of the two components are separated by an angle β determined by Equation 6.

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{w_s \cdot G(\varepsilon_s; \mu_s, \sigma_s) + w_r \cdot G(\varepsilon_s; \mu_r + \beta, \sigma_r) + (1-w_s-w_r) \cdot U(-180, 180)\}, \tag{10}$$

where the weights w_s , w_r and $(1 - w_s - w_r)$ represent the relative contributions (or *intakes*) of spatiotopic processing (with accuracy μ_s and precision $1/\sigma_s$), retinotopic processing (with accuracy μ_r and precision $1/\sigma_r$), and guess rate, respectively. The means of the two Gaussian distributions are separated by an angle β determined by Equation 6.

Models SP3c_S and SP3r_S: Spatiotopic Gaussian + Uniform + Spatiotopic Misbinding Gaussian These models are similar to models F3c and F3r but all components are assumed to be only spatiotopic:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\left\{w_s \cdot G(\varepsilon_s; \mu_s, \sigma_s) + (1-w_s-w_{sm}) \cdot U(-180, 180) + w_{sm} \cdot \sum_{i=1, i \neq t}^T [G(\varepsilon_s; \mu_s + \varepsilon_{i,t}, \sigma_s)]\right\}, \tag{11}$$

where the first two terms represent the same Gaussian and Uniform distributions as in model SP2_S and the third term represents errors stemming from misbinding. The weights w_s , w_{sm} and $(1 - w_s - w_{sm})$ represent the *intake* of spatiotopic processing (with accuracy μ_s and precision $1/\sigma_s$), misbinding rate, and guess rate, respectively. The misbinding term is expected to also have a Gaussian distribution, with the same

Model SP2_S: Spatiotopic Gaussian + Uniform This model has the same form as model F2:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{w_s \cdot G(\varepsilon_s; \mu_s, \sigma_s) + (1-w_s) \cdot U(-180, 180)\}, \tag{8}$$

where the first component is the Gaussian distribution described in model SP1_S, and the second component is a uniform distribution U over the interval (-180, 180), which represents guessing. The weights w_s and $(1 - w_s)$ represent the *intake* of spatiotopic processing and guess rate, respectively.

Model SP2_R: Retinotopic Gaussian + Uniform This model also has two components:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\{w_r \cdot G(\varepsilon_s; \mu_r + \beta, \sigma_r) + (1-w_r) \cdot U(-180, 180)\}, \tag{9}$$

where the first component is the Gaussian distribution described in model SP1_R, and the second component is a uniform distribution U over the interval (-180, 180), which represents guessing. The weights w_r and $(1 - w_r)$ represent the *intake* of retinotopic processing and guess rate, respectively. The angle β is determined by Equation 6.

Model SP2_SR: Spatiotopic Gaussian + Retinotopic Gaussian + Uniform This model combines models SP2_S and SP2_R and is represented by:

standard deviation as the first Gaussian but with the mean shifted from the first Gaussian by the difference $\varepsilon_{i,t}$ between the cued target's and the misbinding object's directions of motion. Similar to models F3c and F3r, models SP3c_S (*closest cued feature*) and SP3r_S (*closest reported feature*) differ in how the selection operator $S_{i=1; i \neq t}^T[\cdot]$ determines the misbinding item from the set of $(T-1)$ noncued objects.

Models SP3c_R and SP3r_R: Retinotopic Gaussian + Uniform + Retinotopic Misbinding Gaussian These models are similar to models F3c and F3r but all components are assumed to be only retinotopic:

$$\text{CDF}(\varepsilon_s) = \text{CDF}\left\{w_r \cdot G(\varepsilon_s; \mu_r + \beta, \sigma_r) + (1-w_r-w_{rm}) \cdot U(-180, 180) + w_{rm} \cdot \sum_{i=1, i \neq t}^T [G(\varepsilon_s; \mu_r + \varepsilon_{i,t} + \beta_i, \sigma_r)]\right\}, \tag{12}$$

where the first two terms represent the same Gaussian and Uniform distributions as in model SP2_R and the third term represents errors stemming from misbinding. The weights w_s , w_{rm} and $(1 - w_r - w_{rm})$ represent the *intake* of retinotopic

processing (with accuracy μ_r and precision $1/\sigma_r$), misbinding rate, and guess rate, respectively. The mean of the first Gaussian is shifted from the origin by an angle β determined by Equation 6.

Models SP3c_SR and SP3r_SR: Spatiotopic Gaussian + Retinotopic Gaussian + Uniform + Spatiotopic Misbinding Gaussian + Retinotopic Misbinding Gaussian
These two models are represented by the following equation:

$$\begin{aligned} \text{CDF}(\varepsilon_s) = & \text{CDF} \left\{ w_s \cdot G(\varepsilon_s; \mu_s, \sigma_s) + w_r \cdot G(\varepsilon_s; \mu_r + \beta, \sigma_r) \right. \\ & \left. + (1 - w_s - w_r - w_{sm} - w_{rm}) \cdot U(-180, 180) \right. \\ & \left. + S_{i=1:i \neq t}^T [w_{sm} \cdot G(\varepsilon_s; \mu_s + \varepsilon_{i,t}, \sigma_s) + w_{rm} \cdot G(\varepsilon_s; \mu_r + \varepsilon_{i,t} + \beta_j, \sigma_r)] \right\}, \end{aligned} \quad (13)$$

where the first three terms are the same spatiotopic Gaussian, retinotopic Gaussian and the Uniform distributions as in model SP2_SR, and the last two terms represent errors stemming from misbinding reports. The selection operator $S_{i=1:i \neq t}^T[\cdot]$ determines from the set of $(T-1)$ noncued objects the misbinding item. Again, this can be either the *closest cued feature* item (model SP3c_SR) or the *closest reported feature* item (model SP3r_SR). Similar to the selected target, this misbinding item also produces a spatiotopic (fourth term) and a retinotopic (fifth term) Gaussian.

Results

Eye-tracking data

In our experiments, observers were required to follow eye-movement instructions while paying sufficient attention to the stimuli. Too many or too frequent invalid fixations/SPEMs in a block might lead to unreliable data because observers would then put most of their effort into the task of gaze control. Therefore, we only accepted blocks with the proportion of valid trials $\geq 50\%$ (see Method section for validity criteria in each condition). That is, a block was excluded if more than 56 trials were needed to obtain 28 valid trials. However, as shown below, we obtained a much higher proportion of valid trials on average. In addition, when an observer had 5 trials rejected in succession during a block of trials, we assumed the eye tracker did not hold calibration, or more likely, the calibration itself was inaccurate, due, presumably, to head or body movements. When this happened part way during a run, we paused the experiment to adjust and recalibrate the eye tracker and then resumed from where the run was paused. We allowed two such interruptions per block. However, we generally discarded the block and gave observers a break if performance did not improve much after each recalibration. Finally, averaging across acceptable blocks, all observers had the proportion of valid trials $>85\%$

in fixation experiments and $>70\%$ in SPEM experiments (less than 824 and 1,000 trials were needed to obtain 700 valid trials, respectively).

Figure 5 plots some examples of two-dimensional gaze traces in the SPEM condition. Also, we projected and computed the changes of eye positions along the pursuit target's direction of motion during each trial. The relative positions between eye and pursuit target are shown in Fig. 6 as a function of time for some other example trials in the SPEM condition. Figures 5 and 6 suggest that, in general, the eye pursued the target with both directional and positional deviations. During the presentation of the stimuli, the eye might move in a direction not perfectly aligned with that of the pursuit target, or slightly lagging behind or running ahead of the target. To ensure accuracy, our approach for the decomposition of spatiotopic and retinotopic components shown in Fig. 4 was therefore performed based on the *actual* eye-velocity vector instead of that of the pursuit target. However, the use of theoretical (i.e., pursuit target) velocity also produced very similar results (see Supplemental Information). For each trial, we fit a line to the last 200-ms part of eye movement trajectory during which the stimuli were presented, and calculated the actual direction of pursuit according to the slope of the line. The magnitude of the eye velocity vector was taken as the mean velocity over the critical period that had been calculated when considering pursuit gain (see Criteria for a valid smooth pursuit trial).

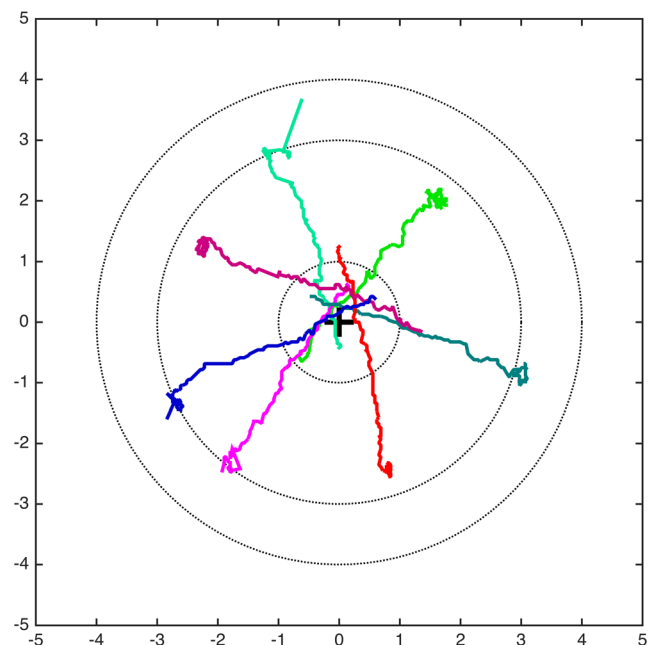


Fig. 5 Two-dimensional gaze traces on example SPEM trials (observer TAN, Experiment 2a) shown in different colors within the 10×10 -deg central area of the screen. The 3-deg, 4-deg, and 1-deg gray circles represent the initial fixation, jump-back, and terminal positions of the pursuit target, respectively. The central cross marks the center of the screen (Color figure online)

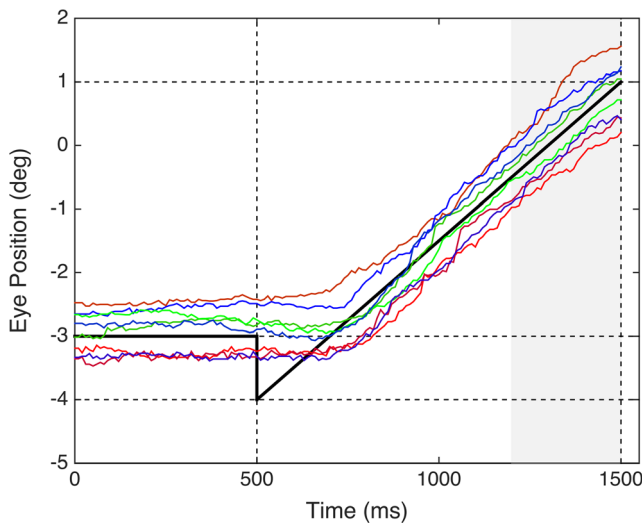


Fig. 6 Eye position (colored lines; seven trials in Experiment 2a, observer TAN) shown with the pursuit target’s position (black solid line) along the pursuit direction as a function of time. The 0-deg position represents the center of the display. The shaded area represents the critical time window within which pursuit gain must fall in the specified range, and no saccades and blinks are allowed (see also Fig. 2) (Color figure online)

In Fig. 7, the green line shows the trace of eye velocity on an example trial. Velocity was computed by digital differentiation of eye position in the direction of target motion after every 10 ms (display sampling frequency = 100 Hz). To reduce noise, we used a low-pass filter with a cutoff frequency of 20 Hz (Butterworth, order = 10). The filtered trace is in blue. As mentioned earlier, averaged velocity (and that for either its horizontal or vertical component, neither of which is shown here) during the last 300-ms period had to fall in the

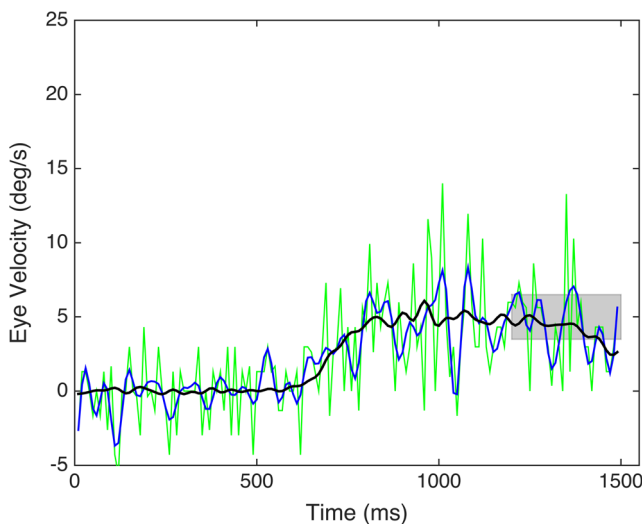


Fig. 7 Eye velocity data for an example SPEM trial (observer TAN, Experiment 2a): raw velocity (green) and low-pass filtered (blue) data. The black line represents the average of filtered traces obtained from 100 randomly selected trials (observer TAN, Experiment 2a). The shaded area shows the constrained range for averaged velocity in the critical pursuit interval (Color figure online)

range [0.7, 1.3] of target velocity (gray shaded area; target velocity = 5 deg/s). The black line shows the average of filtered traces obtained from 100 randomly selected trials. In general, there is a gradual drop of eye velocity in the critical pursuit interval, which can be explained by the observer’s anticipation of when and where the target stops moving (Robinson, Gordon, & Gordon, 1986). However, the requirement for pursuit gain in this interval is still guaranteed.

Overall performance

Experiments 1a and 2a: Stimulus encoding Figure 8 plots error magnitude ($|\varepsilon|$; right y-axis) and transformed performance ($TP = 1 - \frac{|\varepsilon|}{180}$; left y-axis) as a function of set size for the two eye movement conditions: (1) fixation: Experiment 1a, left panel; and (2) SPEM: Experiment 2a, right panel. The transformation metric TP is defined in the same way as in our previous studies (Shooner et al., 2010; Ögmen et al., 2013; Huynh et al., 2015). TP can take on any value in the range [0, 1], in which the values of one and 0.5 correspond to perfect and chance levels of performance, respectively. Because stimulus motion is different according to spatiotopic and retinotopic coordinates in the SPEM condition, we consider SPEM performance measured in each coordinate system separately. We first show here spatiotopic performance ($\varepsilon = \varepsilon_s$ in Fig. 4). A two-way repeated-measures ANOVA with Huynh-Feldt correction for sphericity shows a nonsignificant main effect of *eye movement* (fixation vs. SPEM-spatiotopic) $F(1, 3) = 1.179, p = .357$, a significant main effect of *set size*, $F(3.776, 11.329) = 121.716, p < .0001, \eta_p^2 = 0.976$, and a significant interaction between the two factors, $F(2.832, 8.497) = 4.577, p = .036, \eta_p^2 = 0.604$. This significant interaction appears to be mainly caused by the difference between fixation and SPEM-spatiotopic performance at small set sizes (see Fig. 10). We conducted paired-samples t tests, with Bonferroni correction for multiple comparisons (two-tailed, $\alpha = 0.00714$), to compare fixation and SPEM-spatiotopic performance at different set sizes. A significant difference was found only at a set size of one, $t(3) = 6.568, p = 0.007$. In fact, one observer (DHL) seems to have a different pattern of behavior than the others: This observer’s fixation performance was consistently better than SPEM-spatiotopic performance for *all* set sizes, whereas for other observers fixation performance was consistently better than SPEM-spatiotopic performance for only set size of one. However, we observed no statistical changes when observer DHL was removed from the analyses.

SPEM retinotopic performance was calculated based on the angular deviation between retinotopic and reported motion vectors ($\varepsilon = \varepsilon_r$ in Fig. 4). Retinotopic $|\varepsilon|$ and TP are shown in Fig. 9, left panel. The main effect of *eye movement* becomes significant when comparing fixation with SPEM retinotopic

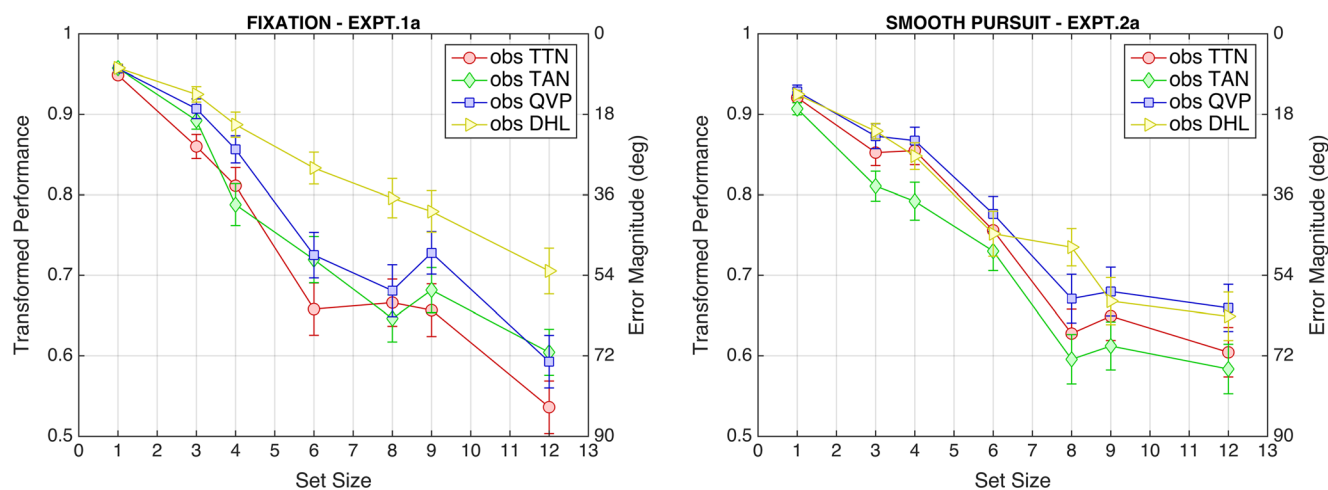


Fig. 8 Data for individual observers in Experiments 1a (fixation; left panel) and 2a (SPEM; right panel): Transformed performance (left y-axis) and error magnitude (right y-axis) plotted as a function of set size.

performance, $F(1, 3) = 16.595$, $p = .027$, $\eta_p^2 = 0.847$, and SPEM spatiotopic with SPEM retinotopic performance, $F(1, 3) = 82.879$, $p = .03$, $\eta_p^2 = 0.965$. In both cases, the main effect of set size and the interaction between eye movement and set size are significant. A one-way repeated-measures ANOVA of retinotopic SPEM performance also returns a significant effect of set size, $F(6, 18) = 20.968$, $p < .0001$, $\eta_p^2 = 0.875$.

Fixation, SPEM-spatiotopic and SPEM-retinotopic performance averaged across observers are shown in Fig. 10. Compared with a similar condition in our previous study (Huynh et al., 2015; Fig. 2, middle panel, blue line), performance observed in both the fixation and SPEM (spatiotopic or retinotopic TP) conditions is worse.² This is predictable because it is likely that, when observers were required to fixate or pursue a target in the present experiment,³ more attention was drawn toward the target and less attention was distributed among the moving stimuli (Intriligator & Cavanagh, 2001). However, the progressive decay of performance with increasing set size, which indicates an early bottleneck of motion processing at the encoding stage, is consistent with our previous findings (Huynh et al., 2015; Öğmen et al., 2013).

Superior performance found during pursuit in spatiotopic coordinates, compared with that in retinotopic coordinates, indicates that spatiotopic encoding dominates and/or has higher precision compared to retinotopic processing. To roughly assess the relative contribution of each component, let us consider the extreme case in which we assume motion is encoded only in a spatiotopic reference frame. If there were no

SPEM performance in the right panel was measured in spatiotopic coordinates. Error bars correspond to ± 1 standard error of the mean (Color figure online)

noise, guessing, or nontarget misreporting in observers' responses, spatiotopic performance is expected to be perfect ($TP = 1.0$) whereas retinotopic performance is at some lower level, which can be calculated based on the average difference between spatiotopic and retinotopic directions of motion (β in Fig. 4). As shown in Fig. 9, right panel, this level of performance is higher than chance (mean ~ 0.75). Because β does not depend on set size or on the actual response of the observers, we observe, as expected, a performance level that is independent of set size, $F(1, 3) = 9.865$, $p = .052$, $\eta_p^2 = 0.767$.⁴ Comparing the left and right panels of Fig. 9, one observes that performance calculated according to the retinotopic reference frame is higher than what one would expect from the case with perfect spatiotopic encoding only for set size one, $t(3) = 10.459$, $p = 0.002$, with $\alpha = 0.007$ with Bonferroni correction). Hence, we can state that for set size one, it is necessary to add the retinotopic reference frame contribution to that of the spatiotopic reference frame to explain the overall performance. On the other hand, we found that performance expressed in terms of a spatiotopic reference frame is not significantly different than overall performance for set sizes three and above. Taken together, our results suggest that, at the stimulus encoding stage, motion stimuli are encoded mainly in a spatiotopic reference frame with a minor contribution from the retinotopic reference frame in the special case of a single target in motion. Inspection of Fig. 10 also shows that

² One observer (DHL) participated in both studies. His performance is only worse in the SPEM condition. There is no clear difference in his performance between the fixation condition and the Huynh et al. study (see Fig. S1.1, middle).

³ In the previous experiment, although we encouraged observers to maintain their eyes at the center of the screen during each trial in order to pay as equal attention as possible to all objects, fixation was not strictly required.

⁴ It is important to note that β is distributed nonuniformly on the interval $[-180, +180]$ across trials. This is because the spatiotopic and retinotopic components are not independent but systematically correlated. The directions of motion for the pursuit target and stimuli were randomly chosen, hence their difference (α in Fig. 4) is random. However, unless α is sufficiently small and eye velocity is bigger than spatiotopic velocity, the value of β is in general smaller than 90 deg (chance level). The same logic holds if one assumes motion is encoded only in retinotopic coordinates. In such a case, perfect responses would produce a retinotopic TP of 1.0 and a spatiotopic TP the same as retinotopic TP (Fig. 9, right panel) when spatiotopic encoding is assumed.

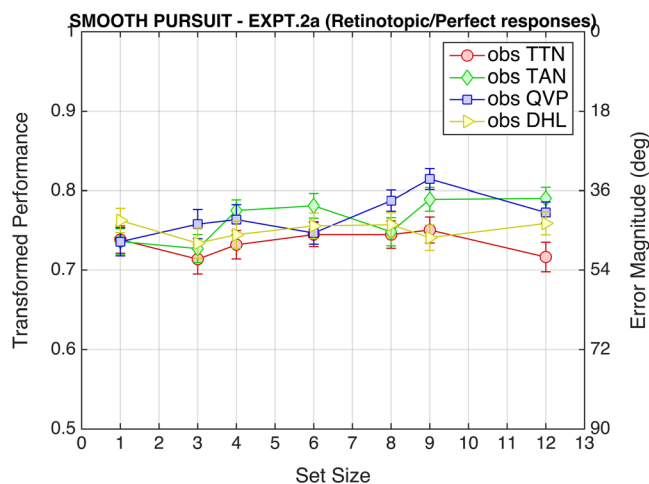
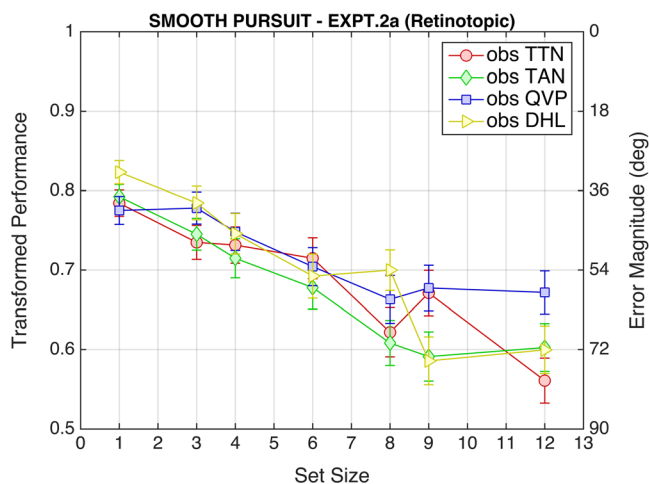


Fig. 9 Data for individual observers in Experiment 2a (SPEM). *Left panel:* SPEM transformed performance (left y-axis) and error magnitude (right y-axis) measured in *retinotopic* coordinates as a function of set size. *Right panel:* same as left panel expressed with

respect to spatiotopic performance, under the assumption that spatiotopic performance is perfect (zero spatiotopic errors). *Error bars* correspond to ± 1 standard error of the mean (Color figure online)

for large set sizes (8–12) the difference between reference frames vanish. Previously, we proposed a leaky-flask model of information processing capacity, which states that significant capacity limits exist prior to memory stages (Huynh et al., 2015; Ögmen et al., 2013). Within the context of this model, we can speculate that at large set sizes, observers start to encode the motion direction of stimuli in more abstract terms such as “moving towards upper right corner” rather than metric encoding in a specific reference frame. Whereas “moving towards upper right corner” may be considered as being based on a spatiotopic reference frame, the key point is that it is a *nonmetric* encoding (there is no explicit quantitative measure of angle). Given that spatiotopic and retinotopic reference

frames are correlated, one would expect the difference between the two reference frames to vanish, even though performance is still better than chance.

Similar observations have been made in previous studies of ensemble coding. Corbett and Melcher (2014) had observers adapting to mean size of dots of various sizes and examined the reference frame used for the resulting size aftereffect. The mean-size aftereffects (a test dot appeared larger following adaptation to small dots and smaller after adaptation to large dots) were seen when the test dot was presented at the appropriate retinotopic or spatiotopic location relative to the adapted region, or even at locations that were neither retinotopic or spatiotopic, but were within the adapted hemifield, suggesting that multiple reference frames are used in the encoding of mean size. Corbett and Melcher suggest that ensemble representations may be available at multiple levels across the hierarchy of visual processing, and these representations efficiently represent abstract, global properties. Even though our study did not explicitly ask observers to report ensemble properties of the stimuli, the use of multiple frames and the abstract encoding of motion may be indicative of the implicit intrusion of ensemble representations for the larger set sizes (see Brady & Alvarez, 2015). However, care must be exercised when extrapolating across feature domains with regard to principles of ensemble coding (Hubert-Wallander & Boynton, 2015).

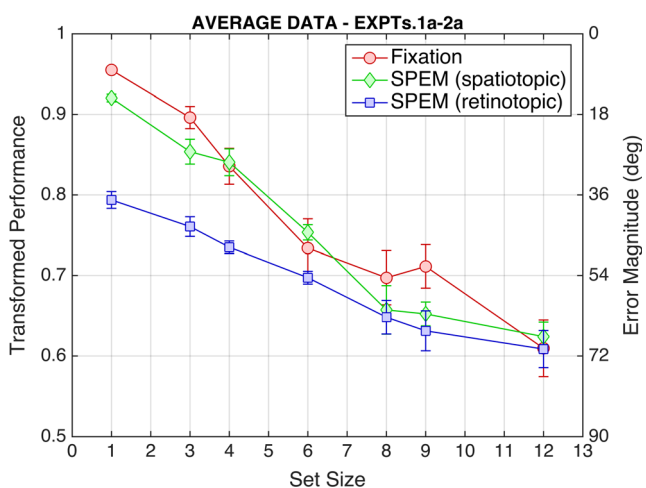


Fig. 10 Average data in Experiments 1a (fixation) and 2a (SPEM): Transformed performance (left y-axis) and error magnitude (right y-axis) averaged across observers as a function of set size for three cases: *Fixation* (red), *SPEM spatiotopic* (green), *SPEM retinotopic* (blue). *Error bars* correspond to ± 1 standard error of the mean (Color figure online)

Experiments 1b and 2b: Sensory memory and VSTM Figure 11 plots fixation (left), SPEM spatiotopic (middle), and SPEM retinotopic (right) performance as a function of cue delay. Average data are shown in Fig. 12. A two-way repeated-measures ANOVA with Huynh-Feldt correction for sphericity shows a significant main effect of *eye movement* (fixation vs. SPEM spatiotopic) $F(1, 3) = 35.583, p = .009, \eta_p^2 = 0.922$, a significant main effect of *cue delay*, $F(6, 18) =$

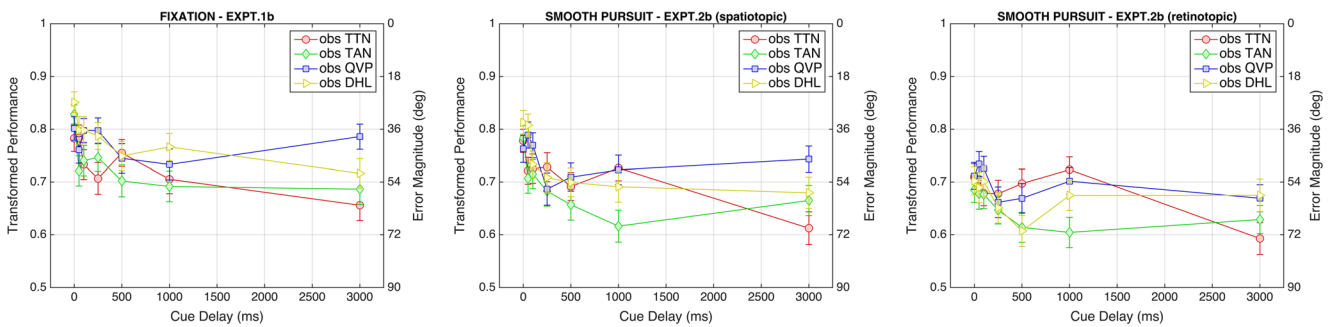


Fig. 11 Data for individual observers in Experiments 1b (fixation; *left panel*) and 2b (SPEM; *center and right panels*): Transformed performance (left y-axis) and error magnitude (right y-axis) plotted as a function of cue delay. Performance during SPEM in the *center and right*

panels was measured in *spatiotopic* and *retinotopic* coordinates, respectively. *Error bars* correspond to ± 1 standard error of the mean (Color figure online)

8.948, $p < .0001$, $\eta_p^2 = 0.749$, and a nonsignificant interaction between the two factors, $F(3.135, 9.405) = 0.762$, $p = .547$, $\eta_p^2 = 0.203$. Given the insignificant difference between fixation and SPEM spatiotopic performance observed at the encoding stage for set size six in Experiments 1a and 2a, we carried out pairwise comparisons to examine whether the significant effect of eye movement we just found exists across all three processing stages (encoding, sensory memory, and VSTM). We grouped the data according to corresponding groups of cue delay samples and ran paired-samples t tests to compare fixation and SPEM spatiotopic TPs for each group. Results from a similar experimental condition in our previous study suggest that the two samples at 1 s and 3 s mainly involve the operation of VSTM, whereas shorter nonzero-delay samples reflect sensory memory (see Huynh et al., 2015; Table 2, row 3). This can be confirmed in the current study by inspection of the average fixation and SPEM spatiotopic performance in Fig. 12, which conform closely to an exponential decay function. Using the same method as in Öğmen et al. (2013) to demarcate sensory and VSTM, we fit observers' average performance in each eye movement condition to an exponential of the form $A + B e^{-t/\tau}$ and obtained time constants (τ) of 292 and 154 ms for the fixation and SPEM (spatiotopic) cases, respectively. Although performance in the latter case reaches steady-state level that represents VSTM earlier ($3\tau = 462$ ms, which precedes the sample at 500 ms), it is more reasonable to keep the demarcation between the two memory systems consistent across eye movement conditions and in the same way as in Huynh et al. (2015). Three paired-samples t -tests (two-tailed, $\alpha = 0.0167$) yield a significant difference between fixation and SPEM spatiotopic TPs at the sensory memory stage, $t(3) = 10.052$, $p = .002$, but not at the encoding, $t(3) = 3.490$, $p = .040$, or VSTM, $t(3) = 3.460$, $p = .041$, stages. The insignificant difference at the encoding stage (zero cue delay) is consistent with the finding in Experiments 1a and 2a. We also find that spatiotopic performance is significantly better than retinotopic performance, $F(1, 3) = 41.217$, $p = .008$, $\eta_p^2 = 0.932$. However, pairwise comparisons (two-tailed, $\alpha =$

0.0167) show that the difference is only significant at the encoding stage, $t(3) = 6.029$, $p = .009$. The insignificant difference found at the two memory stages is not necessarily a hallmark of equivalent contributions of the spatiotopic and retinotopic representations. The drop of spatiotopic performance over time due to increasing guessing and misreporting responses might be the main cause because a one-way repeated-measures ANOVA with Huynh-Feldt correction for sphericity shows that the effect of cue delay on retinotopic performance is not significant, $F(3.051, 9.152) = 3.575$, $p = .059$, $\eta_p^2 = 0.544$. If indeed, at set size six the retinotopic reference frame has no contribution to the performance, as suggested by the findings of Experiment 1, performance plotted in terms of the retinotopic reference frame may represent an overall lower baseline, independent of stimulus encoding and memory stages.

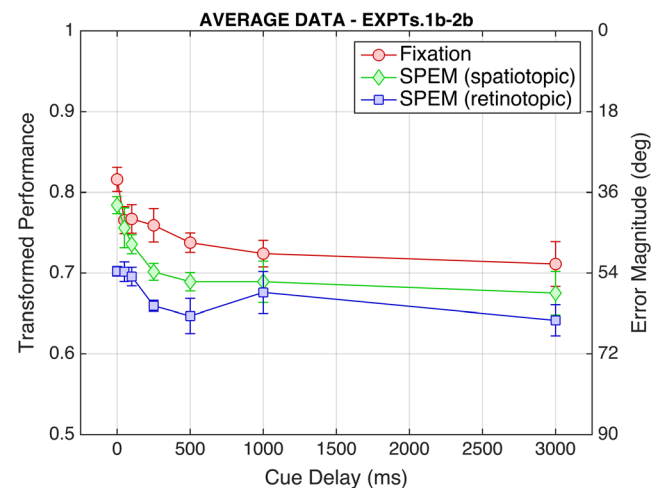


Fig. 12 Average data in Experiments 1b (fixation) and 2b (SPEM): Transformed performance (left y-axis) and error magnitude (right y-axis) averaged across observers as a function of cue delay for three cases: *Fixation* (red), *SPEM spatiotopic* (green), *SPEM retinotopic* (blue). *Error bars* correspond to ± 1 standard error of the mean (Color figure online)

Statistical modeling

Model selection Model selection was used to find the model or group of models that best describe the behavioral data. In the least-squares method, the models were compared based on their adjusted R^2 values,⁵ a measure of their goodness of fit. The model with the highest adjusted R^2 was considered the best-performing model. Table 1 provides average values of adjusted R^2 obtained for all conditions and models. In the fixation condition, our analyses in both Experiments 1a and 1b show that the three models F2, F3c, and F3r have equivalent performance, which is significantly better than that of model F1. Model F2 was selected because it contains the smallest number of free parameters. In the SPEM condition, as stated in the Data Analysis section, our models can be grouped based on two factors (i.e., uncertainty and reference frame). The former one is the same as used to formulate the models in the fixation condition, which is to consider whether a model accounts for guessing and misreporting in the observers' responses. The latter factor indicates the nature of reference systems (spatiotopic, retinotopic, or combined) associated with the encoding and retention of information assumed by each model. Consistently across Experiments 1b and 2b, we find that performance is equivalent for the SP2_*, SP3c_*, and SP3r_* groups, which is significantly better than the SP1 group. Equivalent performance was also found for the spatiotopic (*_S) and combined (*_SR) groups, which is significantly better than the retinotopic (*_R) group. The result does not change when comparing the spatiotopic (*_S) and combined (*_SR) groups for each memory stage separately. Taken together, the model SP2_S was chosen for its smallest number of parameters. This finding implies that a spatiotopic model is sufficient to fully account for the variability in the observers' behavior and reinforces our speculation above that the encoding and retention of motion information is essentially spatiotopic.

Parameter estimation We report in this section estimates for the parameters of the winning model in each condition. Figure 13 plots averaged values for *intake* (along with *guess rate* = $1 - \textit{intake}$) and *precision* obtained in Experiments 1a (fixation) and 2a (SPEM) as a function of set size. As mentioned in the Data Analysis section, *intake* (w in model F2 or w_s in model SP2_S) and *precision* ($1/\sigma$ in model F2 or $1/\sigma_s$ in model SP2_S), respectively, represent the quantitative and qualitative aspects of performance. We observe a linear drop

⁵ The following equation was used to compute adjusted R^2 :

$$\text{Adjusted } R^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$$

where n is the sample size and p is the number of independent variables (parameters) in the model.

of intake with increasing set size in both the fixation and SPEM conditions, and there is no significant difference between the two conditions. The relationship between precision and set size is nonlinear with a big difference between the two conditions at a set size of one. However, this difference in precision gradually vanishes at larger set sizes, which explains the superior fixation performance only at set size of one we found earlier. It should be noted here that we excluded this set size when comparing our models (see Table 1) because models that contain misbinding components are not applicable when there is only a single object presented. For the set size of one, model comparison was run separately, and the winning models remain the same as those for the other cases (model F2 in the fixation condition and the spatiotopic model SP2_S in the SPEM condition). This suggests that the special finding at a set size of one does not come from any apparent influence of retinotopic processing but, presumably, from a higher depletion of attentional resources caused by oculomotor control in the SPEM condition compared with that in the fixation condition. The increase of uncertainty when having larger numbers of objects might have rendered precision in both the fixation and SPEM conditions to drop to a level at which the difference in attentional deployment was no longer noticeable.

Figure 14 plots averaged values for *intake* and *precision* obtained in Experiments 1b (fixation) and 2b (SPEM) as a function of cue delay. Recall that these two experiments consistently used a set size of six while varying cue delay to examine the sensory and VSTM stages of information processing. Given a bottleneck of processing at the encoding stage demonstrated by the degradation of performance with increasing set size in Experiments 1a and 2a, analyzing the extent to which the degradation of performance changes over time provides information about the distribution of information loss across different processing stages. According to our previous findings (Huynh et al., 2015; Öğmen et al., 2013) and preliminary data for the present experiments (not shown), performance at a set size of one is relatively stable over the interval 0–3 s. In Fig. 14, this is represented by the horizontal dashed lines extended from the single data points at zero cue delay (obtained in the single object condition in Experiments 1a and 2a). The pattern of results for both intake and precision is similar in the fixation and SPEM conditions and is consistent with our findings in Öğmen et al. study and with the case of cueing position and reporting direction of motion in Huynh et al. (2015). That is, we find that most of the decay in precision occurs at the encoding stage, whereas the decay in intake is more gradual. A one-way repeated-measures ANOVA of precision shows no effect of cue delay in either the fixation or SPEM conditions. For intake, approximately half of the decay is at the encoding stage. These findings are in agreement with our leaky-flask model proposed in Öğmen et al. (2013).

Table 1 Mean adjusted- R^2 values, obtained from the least-squares fitting method, for different models and conditions

Condition model	Fixation						SPEM									
	F1	F2	F3c	F3r	SP1_S	SP2_S	SP3c_S	SP3r_S	SP1_R	SP2_R	SP3c_R	SP3r_R	SP1_SR	SP2_SR	SP3c_SR	SP3r_SR
Experiment 1a-2a (Set size = N)																
$N = 3$	0.9951	0.9963	0.9963	0.9965	0.9949	0.9963	0.9964	0.9966	0.9056	0.9040	0.9024	0.9024	0.9966	0.9966	0.9966	0.9966
$N = 4$	0.9915	0.9957	0.9956	0.9959	0.9947	0.9969	0.9969	0.9969	0.9091	0.9052	0.9016	0.9016	0.9943	0.9949	0.9949	0.9949
$N = 6$	0.9847	0.9962	0.9965	0.9963	0.9909	0.9952	0.9953	0.9950	0.9657	0.9716	0.9650	0.9650	0.9893	0.9912	0.9917	0.9914
$N = 8$	0.9846	0.9968	0.9972	0.9970	0.9917	0.9972	0.9972	0.9977	0.9892	0.9897	0.9898	0.9898	0.9925	0.9959	0.9963	0.9959
$N = 9$	0.9877	0.9963	0.9966	0.9966	0.9876	0.9963	0.9961	0.9961	0.9858	0.9872	0.9875	0.9886	0.9872	0.9919	0.9917	0.9914
$N = 12$	0.9887	0.9974	0.9974	0.9974	0.9887	0.9965	0.9968	0.9963	0.9875	0.9878	0.9873	0.9873	0.9909	0.9954	0.9955	0.9954
Experiment 1b-2b (Cue delay = D ms)																
$D = 0$	0.9903	0.9961	0.9963	0.9976	0.9897	0.9959	0.9965	0.9962	0.9538	0.9532	0.9417	0.9417	0.9949	0.9959	0.9962	0.9961
$D = 50$	0.9894	0.9972	0.9972	0.9976	0.9875	0.9929	0.9953	0.9930	0.9645	0.9640	0.9633	0.9633	0.9847	0.9872	0.9889	0.9863
$D = 100$	0.9849	0.9962	0.9964	0.9962	0.9875	0.9963	0.9968	0.9971	0.9729	0.9754	0.9730	0.9730	0.9913	0.9939	0.9932	0.9934
$D = 250$	0.9854	0.9966	0.9969	0.9968	0.9888	0.9973	0.9978	0.9974	0.9768	0.9769	0.9769	0.9776	0.9927	0.9952	0.9956	0.9952
$D = 500$	0.9871	0.9967	0.9972	0.9969	0.9898	0.9964	0.9964	0.9972	0.9862	0.9861	0.9860	0.9904	0.9922	0.9950	0.9950	0.9950
$D = 1000$	0.9837	0.9958	0.9959	0.9958	0.9890	0.9958	0.9960	0.9958	0.9836	0.9834	0.9831	0.9838	0.9921	0.9936	0.9943	0.9928
$D = 3000$	0.9865	0.9975	0.9976	0.9978	0.9915	0.9970	0.9967	0.9967	0.9831	0.9829	0.9826	0.9822	0.9874	0.9893	0.9909	0.9892

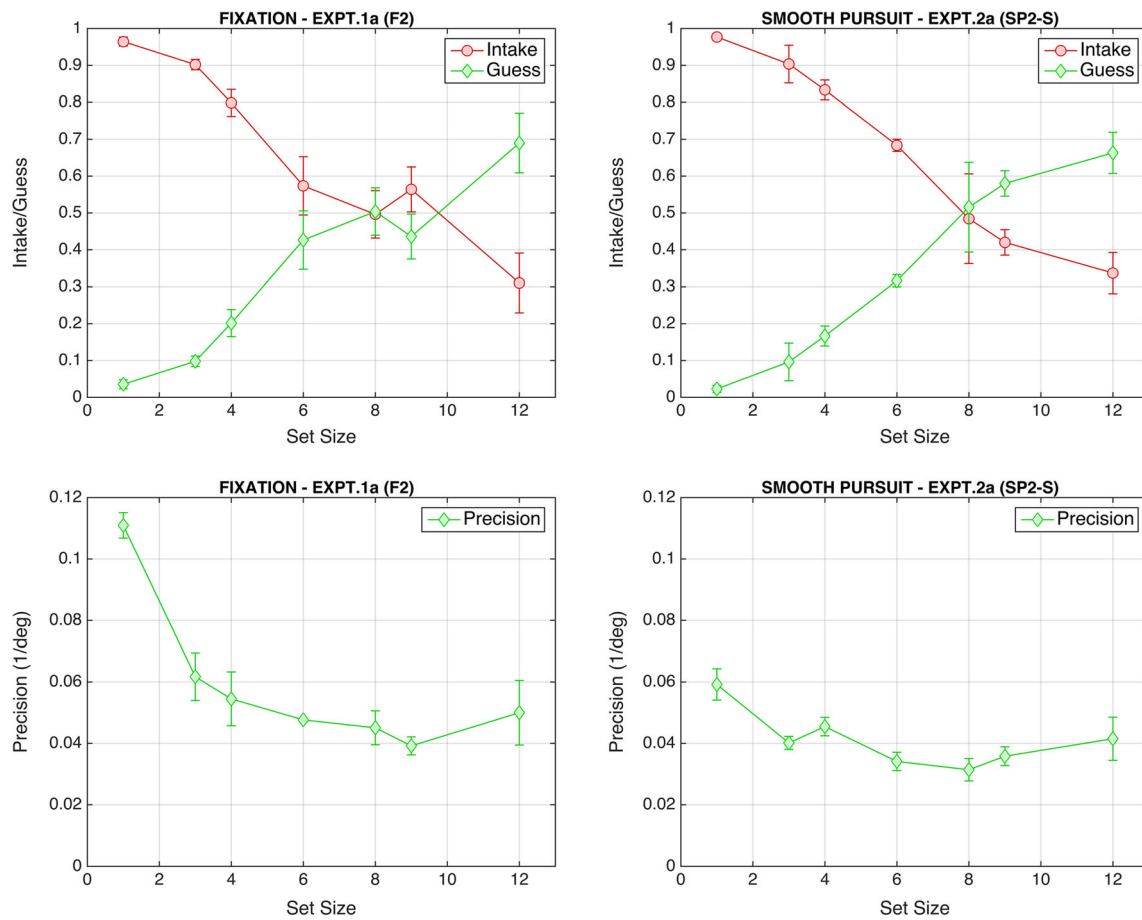


Fig. 13 Decomposition of performance in Experiments 1a (fixation; *left column*) and 2a (SPEM; *right column*): Intake along with guess rate (*upper row*) and precision (*lower row*), averaged across observers, are

shown as a function of set size. Data are shown for only the winning model in each condition (see top of each panel). *Error bars* correspond to ± 1 standard error of the mean (Color figure online)

General discussion

This study aimed to investigate the reference frame used in perceptual encoding and storage of visual motion information. In our experiments, observers viewed multiple moving objects and reported the direction of motion of a randomly selected item. The task was performed while the observers were either fixating a stationary point or smoothly pursuing a target moving at a constant velocity. In the fixation condition, the nonretinotopic component of a motion stimulus is fully confounded with its retinotopic component. In the SPEM condition, with eyes moving from one position to another, the two components can be dissociated. Using a vector decomposition technique, we were able to compute performance during SPEM with respect to spatiotopic (nonretinotopic) and retinotopic motion components and compare them with performance during fixation, which serves as the baseline. We also used several hypothetical models to quantitatively and qualitatively simulate different aspects, including the possible involvement of each reference frame, of the observers' behaviors.

For the stimulus encoding stage, which precedes memory, we found that the reference frame depends on stimulus set size. For the special case where the stimulus consists of a single moving target, the spatiotopic reference frame had the most significant contribution with some additional contribution from the retinotopic reference frame. To a close approximation, the relative contributions of the two reference frames can be quantified based on the two extreme cases we discussed earlier for a set size of one. Average performance at a set size of one in the fixation condition (approximately 0.96; Fig. 8, left panel) can be considered as SPEM spatiotopic performance if motion is assumed to be encoded exclusively in a spatiotopic reference frame. On the other hand, if no spatiotopic reference frame is used, SPEM spatiotopic performance is expected to be about 0.75 (Fig. 9, right panel). There is a total drop of 0.21 in performance between the two extremes. In reality, we obtained a SPEM spatiotopic performance of 0.92. This corresponds to a drop of 0.04, approximately one fifth of the total drop. Therefore, the contribution ratio of spatiotopic to retinotopic reference frames is roughly 4:1. The contribution of both retinotopic and spatiotopic

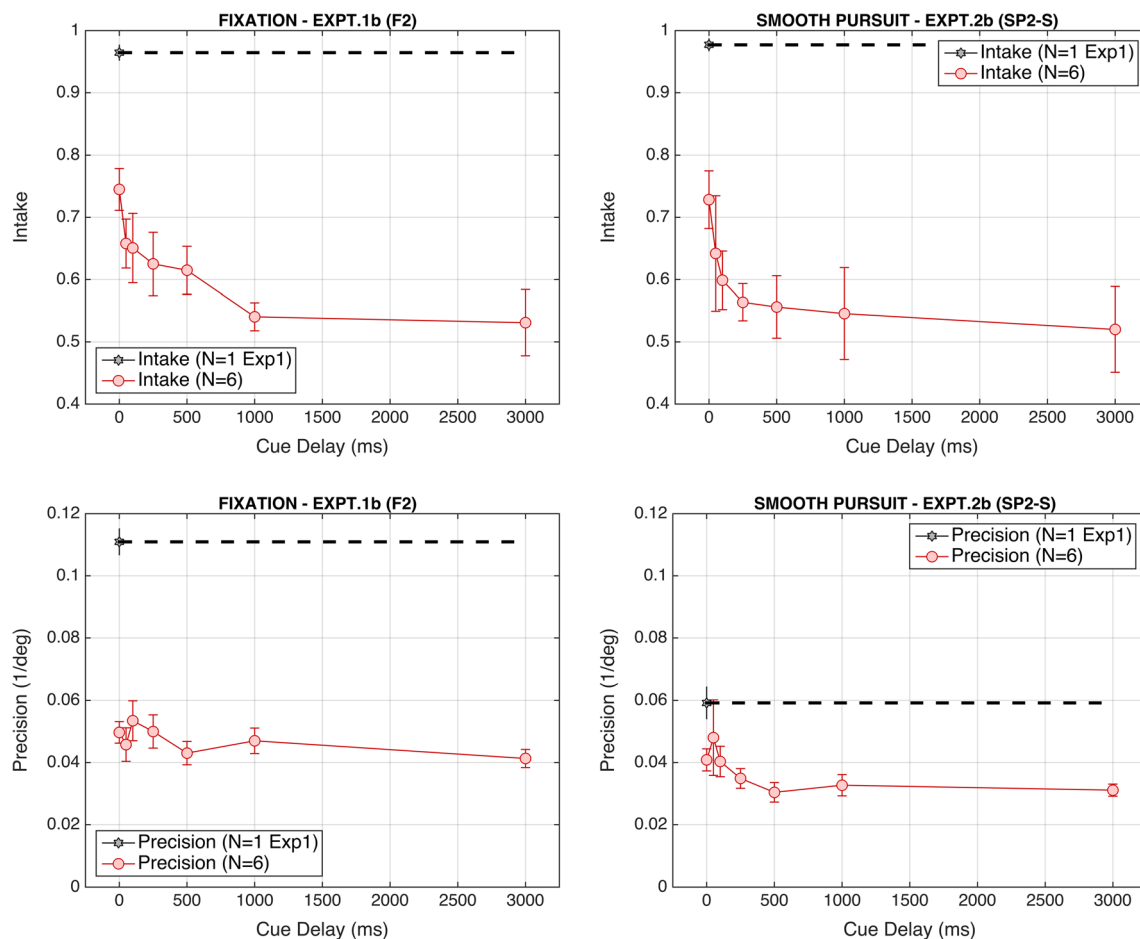


Fig. 14 Decomposition of performance in Experiments 1b (fixation; *left column*) and 2b (SP2-S; *right column*): Intake (*upper row*) and precision (*lower row*), averaged across observers, are shown as a function of cue delay. Data are shown for only the winning model in each condition (see

top of each panel). Error bars correspond to ± 1 standard error of the mean. Data for a set size of 1, shown only at cue delay = 0 s, are taken from Experiment 1. Horizontal lines are to indicate that performance in this condition is largely independent of cue delay (Color figure online)

reference frames for isolated moving targets is in agreement with previous studies (Souman et al., 2005a, 2005b, 2006a; Swanston & Wade, 1988). Although the relative contributions of the reference frames were not provided in these studies, some comparisons can be made between our and their data. To account for errors in motion perception during SPEMs, these studies applied a linear model (Von Holst, 1954) in which the perceived head-centric velocity \mathbf{h}' of a stimulus is viewed as a weighted sum of its retinal image velocity \mathbf{r} and eye velocity \mathbf{e} , that is, $\mathbf{h}' = \rho \cdot \mathbf{r} + \varepsilon \cdot \mathbf{e}$ (for alternative models, see Freeman, 2001; Turano & Massof, 2001; Wertheim, 1994). To compute \mathbf{h}' , the visual system obtains estimates of the actual signals \mathbf{r} and \mathbf{e} . The weights ρ and ε in the model describes the gains associated with these estimates. The deviation of the perceived direction \mathbf{h}' from the physical direction \mathbf{h} depends on the gain ratio ε/ρ . During SPEM, the direction of \mathbf{h}' is typically biased towards the direction of \mathbf{r} , which can be explained by a gain ratio that is smaller than one. The smaller the gain ratio is, the larger the bias becomes. In case that $\varepsilon = \rho = 1$, $\mathbf{h}' = \mathbf{h} = \mathbf{r} + \mathbf{e}$. For example, Souman et al. (2005b)

measured the perceived motion direction of a stimulus moving at various angles (0° – 360°) relative to the pursuit direction. The perceived direction data were fit to the linear model above with the gain ratio ε/ρ being the only free parameter, which was assumed to be fixed across stimulus directions. Souman et al. found a high degree of fit ($R^2 \sim 90\%$) for most observers. They obtained relatively low estimates for ε/ρ , and this ratio decreased with increasing stimulus speed ($3^\circ/\text{s}$: mean = 0.53, standard deviation = 0.12; $8^\circ/\text{s}$: mean = 0.21, standard deviation = 0.1; calculations are based on data in Table 1, Souman et al., 2005b). One can predict that, if the same stimulus speed as in our study ($5^\circ/\text{s}$) were used, the mean gain ratio would be smaller than 0.53 and greater than 0.21. For comparison, we applied the same linear model and simulation on our data for the set size of one in Experiment 2a and obtained a value of ε/ρ that is much higher than predicted ($\varepsilon/\rho = 0.80, 0.63, 0.77, 0.63$ for observers TTN, TAN, QVP, DHL, respectively; mean = 0.71, standard deviation = 0.09). This suggests that our observers generally made smaller errors in judging the direction of motion of the stimulus, and one can conclude that the

data in the Souman et al. study show a larger contribution of the retinotopic reference frame compared to our finding. Let us note, however, that the estimation of contribution ratio we obtained earlier is only meaningful on average data (i.e., performance for different deviations between the pursuit target and the direction of stimulus motion is averaged). The reason for this is that, given that the visual system uses some fixed gain ratio for different stimulus directions, the magnitude of errors of the judged motion direction during SPEM typically depends on the angle between the stimulus and pursuit target motion directions. With the exception of Souman et al. (2005b), most previous studies of motion perception during SPEM (Becklen et al., 1984; Souman et al., 2005a, 2006a, 2006b; Swanston & Wade, 1988; Wallach et al., 1985) focus on horizontal and vertical movements of stimuli and the pursuit target. Therefore, it is hard to quantitatively compare those with our study. However, one potentially important factor that might amplify the contribution of retinotopic encoding in all of these studies is that their experiments were performed in total darkness, with only the stimulus and the pursuit target visible. This would have eliminated the stationary background and the display as usable spatiotopic reference frames. The use of a relatively short stimulus presentation duration (200 ms) in our experiments is unlikely to be a reason for the weak effect of retinotopic reference frame we observed. It has been shown that decreasing the stimulus presentation duration increases errors (biases) in the perceived motion direction during pursuit (De Graaf & Wertheim 1988; Mack & Herman, 1978; Souman et al., 2005a). Therefore, the shorter the presentation duration, the stronger the expected contribution of the retinotopic reference frame. Furthermore, as shown in Souman et al. (2005a), the effect of stimulus duration is negligible for low stimulus velocities, such as that used in our experiments ($5^\circ/\text{s}$).

When the number of items in the stimulus increased, the spatiotopic reference frame alone was able to account for the overall performance. Finally, when the number of items became large, the distinction between reference frames vanished. We interpret this finding as a switch to a more abstract encoding of motion direction, such as “towards lower right” instead of a metric encoding within a specific reference frame. Our earlier studies showed significant capacity limits already at the stimulus encoding stage, leading to the leaky flask model (see Ögmen et al., 2013; Fig. 10). The results of this study are also in agreement with the leaky flask model. When the stimulus set size increases, due to capacity limits, it may not be possible for the visual system to encode all directions of motion according to a reference frame metric. One strategy would be then to switch to a more descriptive nonmetric encoding. Another way the visual system can handle the complexity of a stimulus comprising multiple moving targets (large set size) is through Gestalt grouping mechanisms. For example, the point lights placed on a person in the biological motion paradigm (Johansson, 1973) creates a very

complex stimulus; however, by grouping these points into a meaningful Gestalt (Yantis, 1992), the visual system is capable of computing a common reference frame, which can be used to simplify the *relative* motions of various point lights. Several studies showed that when the stimulus allows grouping of parts, motion groupings based nonretinotopic reference frames (relative motion) account for perceived direction of motion (Agaoglu et al., 2015a, 2015b; Boi, Ögmen, Krummenacher, Otto, & Herzog, 2009; Duncker, 1929/1938; Johansson, 1973; Noory et al., 2015). In fact, Agaoglu et al. (2015b) quantified the contributions of retinotopic, spatiotopic, and relative-motion reference frames and showed that relative motion dominated both during fixation and SPEM, with a contribution more than 80% when the distance between the stimuli was 2 degrees. The dominance of the relative motion decreased with the distance between stimuli; however, for separations as large as 11 deg, the contribution of relative motion was still substantial (60%). Each disk in our experiments here had an independently and randomly chosen direction and hence our stimulus was not conducive to this type of (relative) nonretinotopic reference frame. Instead, the nonretinotopic reference frame was presumably a screen-based (spatiotopic) reference frame.

SPEM not only causes biases in motion perception but also leads to mislocalizations of stimuli (Brenner, Smeets, & Van den Berg, 2001; Mita, Hironaka, & Koike, 1950; Rotman, Brenner, & Smeets, 2005; Souman et al., 2006a; Van Beers, Wolpert, & Haggard, 2001; Ward, 1976). Importantly, both the magnitude and the direction of mislocalization when pursuit is towards the stimulus are different from those when pursuit is away from the stimulus (Mateeff & Hohsbein, 1988; Mateeff et al., 1991; Mitrani & Dimitrov, 1982; Rotman, Brenner, & Smeets, 2004). Also, mislocalization is more pronounced for stimuli that are more distant from the pursuit path (Souman et al., 2006a). We expect that these asymmetric effects on localization also apply to direction-of-motion perception, although the direct physical relationship between location and motion may not exist in perceptual terms (Snowden, 1994; Souman et al., 2006a). It should be noted that, because the cued target’s location and motion direction, as well as the direction of pursuit were chosen randomly across trials, these asymmetries should be averaged out and are not considered in our data.

In terms of memory, we found that performance expressed in terms of a retinotopic reference frame did not depend on cue delay, suggesting that the retinotopic reference frame was not used during memory storage. The difference between fixation and SPEM performance in terms of a spatiotopic reference frame was not significant at stimulus encoding and VSTM stages, but it was significant for the sensory memory stage. As mentioned before, whereas earlier studies found that sensory memory uses a retinotopic reference frame, more recent studies using sequential metacontrast and Ternus-Pikler displays indicate that sensory memory can also use a motion-based nonretinotopic reference frame (Noory et al., 2015; Ögmen et al., 2006; Otto, Ögmen, &

Herzog, 2006; Scharnowski, Hermens, Kammer, Ögmen, & Herzog, 2007). Because performance in sensory memory in terms of a spatiotopic reference frame was superior to that in terms of a retinotopic reference frame and because spatiotopic performance showed the typical exponential delay observed in sensory memory (see Fig. 12), we conclude that our findings here are in agreement with the existence of a nonretinotopic sensory memory component.

When set size was large (8–12 objects) motion performance expressed in a spatiotopic reference frame in the SPEM condition was comparable to that expressed in a retinotopic reference frame. Similar findings have been reported in a study that investigated ensemble coding of mean size using adaption aftereffects and found that adaptation to mean size occurred in multiple reference frames, that included retinotopic and spatiotopic frames, among others (Corbett & Melcher, 2014). The study also reported an abstract representation of mean size, similar to the abstract representation of motion direction we find in the current study. While the current study did not specifically investigate ensemble encoding of motion, the influence of ensemble coding, for the larger set sizes we investigated, cannot be ruled out (see Brady & Alvarez, 2015). Corbett and Melcher (2014) interpreted their findings to suggest that mean size is represented in multiple levels in the visual hierarchy, and this is important for perceptual stability. It is likely that, for large set sizes, motion direction too is represented in multiple reference frames at different levels in the visual hierarchy, facilitating perceptual stability. However, how principles of ensemble coding in one perceptual task generalize to another is an open question (Hubert-Wallander & Boynton, 2015).

In summary, our results, along with other recent findings, suggest that, whereas a retinotopic reference frame may be useful for controlling eye-movements, nonretinotopic reference frames may characterize perception and memory. Furthermore, the use of a nonretinotopic reference frame appears to be capacity limited. In the case of complex stimuli (large set size), the visual system may use perceptual grouping or summary statistics or ensemble representations in order to simplify the complexity of stimuli (as in studies involving mean-size, biological motion, or Ternus-Pikler displays), or resort to a nonmetric abstract coding of motion information.

References

- Agaoglu, M. N., Herzog, M. H., & Ögmen, H. (2015a). Field-like interactions between motion-based reference frames. *Attention, Perception & Psychophysics*, *77*, 2082–2097.
- Agaoglu, M. N., Herzog, M. H., & Ögmen, H. (2015b). The effective reference frame in perceptual judgments of motion direction. *Vision Research*, *107*, 101–112.
- Andersen, R. A., Snyder, L. H., Li, C. S., & Stricanne, B. (1993). Coordinate transformations in the representation of spatial information. *Current Opinion in Neurobiology*, *3*, 171–176.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation (Volume 2)* (pp. 89–195). New York: Academic Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.
- Baker, J. T., Harper, T. M., & Snyder, L. H. (2003). Spatial memory following shifts of gaze: I. Saccades to memorized world-fixed and gaze-fixed targets. *Journal of Neurophysiology*, *89*, 2564–2576.
- Bauer, B. (2015). A selective summary of visual averaging research and issues up to 2000. *Journal of Vision*, *15*(4), 1–15.
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *10*(7), 1–11.
- Becklen, R., Wallach, H., & Nitzberg, D. (1984). A limitation of position constancy. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 713–723.
- Boi, M., Ögmen, H., Kruppenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. nonretinotopic processing. *Journal of Vision*, *9*(13), 1–11.
- Burr, D. (1980). Motion smear. *Nature*, *284*, 164–165.
- Brady, T. M., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 921–929.
- Braun, D. I., Schütz, A. C., & Gegenfurtner, K. R. (2010). Localization of speed differences of context stimuli during fixation and smooth pursuit eye movements. *Vision Research*, *50*, 2740–2749.
- Bremner, A. J., Bryant, P. E., & Mareschal, D. (2005). Object-centred spatial reference in 4-month-old infants. *Infant Behaviour and Development*, *29*, 1–10.
- Brenner, E., Smeets, J., & Van den Berg, A. V. (2001). Smooth eye movements and spatial localisation. *Vision Research*, *41*, 2253–2259.
- Bridgeman, B. (1995). A review of the role of efference copy in sensory and oculomotor control systems. *Annals of Biomedical Engineering*, *23*, 409–422.
- Bridgeman, B., Van der Heijden, A. H., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, *17*, 247–258.
- Burr, D. C., & Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society of London B*, *366*(1564), 504–515.
- Burr, D. C., & Morrone, M. C. (2012). Constructing stable spatial maps of the world. *Perception*, *41*(11), 1355–1372.
- Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.
- Cant, J. S., Sun, S. Z., & Xu, Y. (2015). Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. *Journal of Vision*, *15*(4), 1–21.
- Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153.
- Chen, S., Bedell, H. E., & Ögmen, H. (1995). A target in real motion appears blurred in the absence of other proximal moving targets. *Vision Research*, *35*, 2315–2328.
- Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: Mean size is represented across multiple reference frames. *Attention, Perception, & Psychophysics*, *76*, 746–758.
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*, 1716–1722.

- De Graaf, B., & Wertheim, A. H. (1988). The perception of object-motion during smooth-pursuit eye movements: Adjacency is not a factor contributing to the Filehne illusion. *Vision Research*, *28*, 497–502.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*(1), 1–38.
- Duncker, K. (1938). *Induced motion*. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology*. London, UK: Routledge & Kegan Paul. (Original work published in German, 1929).
- Engel, S. A. (1994). fMRI of human visual cortex. *Nature*, *370*(6485), 106–106.
- Festinger, L., Sedgwick, H. A., & Holtzman, J. D. (1976). Visual perception during smooth pursuit eye movement. *Vision Research*, *16*, 1377–1386.
- Freeman, T. C. A. (2001). Transducer models of head-centred motion perception. *Vision Research*, *41*, 2741–2755.
- Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience*, *28*(15), 3988–3999.
- Gibaldi, A., Canessa, A., & Sabatini, S. P. (2016). Visuomotor behavior of dominant and nondominant eye in 3D visual exploration. *ECVP Abstracts*.
- Golomb, J. D., & Kanwisher, N. (2012). Retinotopic memory is more precise than spatiotopic memory. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1796–1801.
- Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences*, *6*, 1–54.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, R751–R753.
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics*, *72*, 1765–1775.
- Howe, P. D. L., Pinto, Y., & Horowitz, T. S. (2010). The coordinate systems used in visual tracking. *Vision Research*, *50*, 2375–2380.
- Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries over time. *Journal of Vision*, *15*(4), 1–12.
- Huynh, D., Tripathy, S. P., Bedell, H. E., & Ögmen, H. (2015). Stream specificity and asymmetries in feature binding and content-addressable access in visual encoding and memory. *Journal of Vision*, *15*(13), 14, 1–32.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216.
- Irwin, D. E., Brown, J. S., & Sun, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology*, *117*, 276–287.
- Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, *34*, 49–57.
- Jancke, D., Chavane, F., Naaman, S., & Grinvald, A. (2004). Imaging cortical correlates of illusion in early visual cortex. *Nature*, *428*, 423–426.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, *14*, 201–211.
- Jonides, J., Irwin, D. E., & Yantis, S. (1983). Failure to integrate information from successive fixations. *Science*, *222*, 188.
- Levi, D. M., & Tripathy, S. P. (2006). Is the ability to identify deviations in multiple trajectories compromised by amblyopia? *Journal of Vision*, *6*, 1367–1379.
- Mack, A. (1986). Perceptual aspects of motion in the frontal plane. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. I, pp. 1–38). New York: Wiley.
- Mack, A., & Herman, E. (1978). The loss of position constancy during pursuit eye movements. *Vision Research*, *18*, 55–62.
- Mateeff, S. (1980). Visual perception of movement patterns during smooth eye tracking. *Acta Physiologica et Pharmacologica Bulgarica*, *6*, 82–89.
- Mateeff, S., & Hohnsbein, J. (1988). Perceptual latencies are shorter for motion towards the fovea than for motion away. *Vision Research*, *28*, 711–719.
- Mateeff, S., Yakimoff, N., Hohnsbein, J., Ehrenstein, W. H., Bohdanecy, Z., & Radil, T. (1991). Selective directional sensitivity in visual motion perception. *Vision Research*, *31*, 131–138.
- Mays, L. E., & Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, *43*, 207–232.
- McKenzie, A., & Lisberger, S. G. (1986). Properties of signals that determine the amplitude and direction of saccadic eye movements in monkeys. *Journal of Neurophysiology*, *56*, 196–207.
- Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences*, *12*, 466–473.
- Melcher, D., & Fracasso, A. (2012). Remapping of the line motion illusion across eye movements. *Experimental Brain Research*, *218*, 503–514.
- Melcher, D., & Morrone, M. C. (2015). Nonretinotopic visual processing in the brain. *Visual Neuroscience*, *32*, E017.
- Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting: Ocular dominance demonstrated by unconscious sighting. *Journal of Experimental Psychology*, *12*, 113–126.
- Miles, W. R. (1930). Ocular dominance in human adults. *Journal of General Psychology*, *3*, 412–429.
- Mita, T., Hironaka, K., & Koike, I. (1950). The influence of retinal adaptation and location on the “Empfindungszeit.”. *Tohoku Journal of Experimental Medicine*, *52*, 397–405.
- Mitrani, L., & Dimitrov, G. (1982). Retinal location and visual localization during pursuit eye movement. *Vision Research*, *22*, 1047–1051.
- Narasimhan, S., Tripathy, S. P., & Barrett, B. T. (2009). Loss of positional information when tracking multiple moving dots: The role of visual memory. *Vision Research*, *49*, 10–27.
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology*, *14*, 830–839.
- Noory, B., Herzog, M. H., & Ögmen, H. (2015). Spatial properties of nonretinotopic reference frames in human vision. *Vision Research*, *113*, 44–54.
- Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, *15*(4), 1–14.
- Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*, 272–288.
- Ögmen, H. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and nonretinotopic representations. *Advances in Cognitive Psychology*, *3*, 67–84.
- Ögmen, H., Ekiz, O., Huynh, D., Tripathy, S. P., & Bedell, H. E. (2013). Bottlenecks of motion processing during a visual glance: The leaky flask model. *PLOS ONE*. doi:10.1371/journal.pone.0083671
- Ögmen, H., & Herzog, M. H. (2010). The geometry of visual perception: Retinotopic and nonretinotopic representations in the human visual system. *Proceedings of the IEEE*, *98*, 479–492.
- Ögmen, H., Otto, T., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, *46*, 3234–3242.
- Ong, W. S., Hooshvar, N., Zhang, M., & Bisley, J. W. (2009). Psychophysical evidence for spatiotopic processing in area MT in a short-term memory for motion task. *Journal of Neurophysiology*, *102*, 2435–2440.
- Orban de Xivry, J. J., & Lefèvre, P. (2007). Saccades and pursuit: Two outcomes of a single sensorimotor process. *Journal of Physiology*, *584*, 11–23.

- Otto, T., Ögmen, H., & Herzog, M. H. (2006). The flight path of the phoenix: The visible trace of invisible elements in human vision. *Journal of Vision*, *6*, 1079–1086.
- Pertzov, Y., Avidan, G., & Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *Journal of Neuroscience*, *31*, 1059–1068.
- Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology*, *159*, 326–338.
- Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics*, *34*, 39–48.
- Robinson, D. A., Gordon, J. L., & Gordon, S. E. (1986). A model of the smooth pursuit eye movement system. *Biological Cybernetics*, *55*, 43–57.
- Rotman, G., Brenner, E., & Smeets, J. B. J. (2004). Quickly tapping targets that are flashed during smooth pursuit reveals perceptual mislocalizations. *Experimental Brain Research*, *156*, 409–414.
- Rotman, G., Brenner, E., & Smeets, J. B. J. (2005). Flashes are localized as if they were moving with the eyes. *Vision Research*, *45*, 355–364.
- Scharnowski, F., Hermens, F., Kammer, T., Ögmen, H., & Herzog, M. H. (2007). Feature fusion reveals slow and fast memories. *Journal of Cognitive Neuroscience*, *19*, 632–641.
- Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350–1354.
- Shooner, C., Tripathy, S., Bedell, H., & Ögmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision*, *10*(6), 1–20.
- Snowden, R. J. (1994). Motion processing in the primate cerebral cortex. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 51–83). London: Academic.
- Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2005a). Vertical object motion during horizontal ocular pursuit: Compensation for eye movements increases with presentation duration. *Vision Research*, *45*, 845–853.
- Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2005b). Perceived motion direction during smooth pursuit eye movements. *Experimental Brain Research*, *164*, 376–386.
- Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006a). Localization and motion perception during smooth pursuit eye movement. *Experimental Brain Research*, *171*, 448–458.
- Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006b). Frame of reference transformations in motion perception during smooth pursuit eye movements. *Journal of Computational Neuroscience*, *20*, 61–76.
- Sun, J. S., & Irwin, D. E. (1987). Retinal masking during pursuit eye movements: implications for spatiotopic visual persistence. *Journal of Experimental Psychology*, *13*, 140–145.
- Swanston, M. T., & Wade, N. J. (1988). The perception of visual motion during movements of the eyes and of the head. *Perception & Psychophysics*, *43*, 559–566.
- Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 329–337.
- Tootell, R. B. H., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 811–817.
- Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., & Billeveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic-resonance-imaging. *Journal of Neuroscience*, *15*(4), 3215–3230.
- Tripathy, S. P., & Barrett, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision*, *4*(12), 1020–1043.
- Tripathy, S. P., & Levi, D. M. (2008). On the effective number of tracked trajectories in amblyopic human vision. *Journal of Vision*, *8*(4), 1–22. doi:10.1167/7.6.2
- Tripathy, S. P., Narasimhan, S., & Barrett, B. T. (2007). On the effective number of tracked trajectories in normal human vision. *Journal of Vision*, *7*(6), 1–18. doi:10.1167/7.6.2
- Turano, K. A., & Massof, R. W. (2001). Nonlinear contribution of eye velocity to motion perception. *Vision Research*, *41*, 385–395.
- Van Beers, R. J., Wolpert, D. M., & Haggard, P. (2001). Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *Journal of Neurophysiology*, *85*, 1914–1922.
- Vikesdal, G. H., & Langaas, T. (2016). Saccade latency and fixation stability: Repeatability and reliability. *Journal of Eye Movement Research*, *9*, 1–13.
- Von Helmholtz, H. (1925). *Treatise on physiological optics* (Vol. 3). New York: Optical Society of America.
- Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*, *2*, 89–94.
- Wade, N. J., & Swanston, M. T. (1996). A general model for the perception of space and motion. *Perception*, *25*, 187–194.
- Wallach, H., Becklen, R., & Nitzberg, D. (1985). The perception of motion during collinear eye movements. *Perception & Psychophysics*, *38*, 18–22.
- Ward, F. (1976). Pursuit eye movements and visual localization. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes*. New York: Wiley.
- Wertheim, A. H. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, *17*, 293–355.
- Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, *48*(20), 2070–2089.
- Yantis, S. (1992). Multi-element visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zelinsky, G., & Todor, A. (2010). The role of “rescue saccades” in tracking objects through occlusions. *Journal of Vision*, *10*(7), 132.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.