Systems/Circuits

Neurons in Primate Area MSTd Signal Eye Movement Direction Inferred from Dynamic Perspective Cues in Optic Flow

Grace F. DiRisio,[1,4] Yongsoo Ra,[1,2] Yinghui Qiu,[1,3] Akiyuki Anzai,[1] and Gregory C. DeAngelis[1]
[1]Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627, [2]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, [3]College of Veterinary Medicine, Cornell University, Ithaca, New York 14853-6401, and [4]Department of Neurobiology, University of Chicago, Chicago, Illinois 60637

Smooth eye movements are common during natural viewing; we frequently rotate our eyes to track moving objects or to maintain fixation on an object during self-movement. Reliable information about smooth eye movements is crucial to various neural computations, such as estimating heading from optic flow or judging depth from motion parallax. While it is well established that extraretinal signals (e.g., efference copies of motor commands) carry critical information about eye velocity, the rotational optic flow field produced by eye rotations also carries valuable information. Although previous work has shown that dynamic perspective cues in optic flow can be used in computations that require estimates of eye velocity, it has remained unclear where and how the brain processes these visual cues and how they are integrated with extraretinal signals regarding eye rotation. We examined how neurons in the dorsal region of the medial superior temporal area (MSTd) of two male rhesus monkeys represent the direction of smooth pursuit eye movements based on both visual cues (dynamic perspective) and extraretinal signals. We find that most MSTd neurons have matched preferences for the direction of eye rotation based on visual and extraretinal signals. Moreover, neural responses to combinations of these signals are well predicted by a weighted linear summation model. These findings demonstrate a neural substrate for representing the velocity of smooth eye movements based on rotational optic flow and establish area MSTd as a key node for integrating visual and extraretinal signals into a more generalized representation of smooth eye movements.

Key words: **eye movements; motion processing; optic flow; smooth pursuit; visual cortex**

Significance Statement

We frequently rotate our eyes to smoothly track objects of interest during self-motion. Information about eye velocity is crucial for a variety of computations performed by the brain, including depth perception and heading perception. Traditionally, information about eye rotation has been thought to arise mainly from extraretinal signals, such as efference copies of motor commands. Previous work shows that eye velocity can also be inferred from rotational optic flow that accompanies smooth eye movements, but the neural origins of these visual signals about eye rotation have remained unknown. We demonstrate that macaque neurons signal the direction of smooth eye rotation based on visual signals, and that they integrate both visual and extraretinal signals regarding eye rotation in a congruent fashion.

Introduction

During natural viewing, we rotate and reorient our body, head, and eyes to focus on relevant objects and collect information about the visual scene. When the eyes rotate smoothly to track a target of interest, there are associated motor signals (i.e., efference copies of motor commands), and there are also visual consequences in the form of global optic flow. Traditionally, these visual consequences have been considered undesirable, thus requiring compensation by extraretinal signals such as efference copies of motor commands (von Holst and Mittelstaedt, 1950; Wallach, 1987; Royden et al., 1992; Bradley et al., 1996; Ben Hamed et al., 2003). However, theoretical work has emphasized

Received Oct. 5, 2022; revised Jan. 18, 2023; accepted Jan. 24, 2023.

Author contributions: A.A. and G.C.D. designed research; G.F.D., Y.R., and Y.Q. performed research; A.A. contributed unpublished reagents/analytic tools; G.F.D., Y.R., Y.Q., and G.C.D. analyzed data; G.F.D. and G.C.D. wrote the paper.

This work was supported by National Eye Institute (NEI) Grant R01-EY-013644 and by Core Grant EY-001319 from the NEI. We thank Johnny Wen and Amanda Yung for assistance with the programming of visual stimuli. We also thank Dina Graf and Emily Murphy for assistance with animal training and care. In addition, we thank Marjena Popovic for assistance with data collection for one animal.

The authors declare no competing financial interests.

Correspondence should be addressed to Gregory C. DeAngelis at gdeangelis@ur.rochester.edu.

https://doi.org/10.1523/JNEUROSCI.1885-22.2023

Copyright © 2023 the authors

that there is useful information in the visual motion that results from eye rotation. Eye rotation produces patterns of optic flow that are distinct from eye translation, such that it is possible to estimate both translational and rotational components of eye movement from optic flow (Longuet-Higgins and Prazdny, 1980; Rieger and Lawton, 1985; Royden, 1997). More recently, several studies have shown that the visual system can perform computations involving eye velocity by making use of this rotational pattern of optic flow, which has also been referred to as "dynamic perspective" cues (Koenderink and van Doorn, 1976; Grigo and Lappe, 1999; Kim et al., 2015; Sunkara et al., 2015; Danz et al., 2020).

Consider the viewing context depicted in Figure 1, in which the observer's eye translates rightward while counterrotating leftward to maintain fixation on a world-fixed target (red cross). This results in a rotation of the checkerboard relative to the axis of gaze, which induces a trapezoidal image distortion under planar image projection (Fig. 1, bottom). Assuming a stationary world, this dynamic perspective distortion could be used to infer the rotation of the eye.

Two main lines of evidence have shown that the brain extracts eye rotation information from rotational optic flow. First, perception of depth from motion parallax requires signals about eye velocity relative to the scene (Nawrot, 2003; Nadler et al., 2009; Nawrot and Stroyan, 2009). These signals have traditionally been attributed to efference copies of motor commands (Nawrot, 2003; Naji and Freeman, 2004; Nadler et al., 2009). However, recent work showed that rotational optic flow was sufficient to generate neural selectivity for depth based on motion parallax in the absence of physical eye movements (Kim et al., 2015). Thus, rotational optic flow can substitute for efference copy in computing depth from motion parallax. Second, perception of heading during pursuit eye movements also requires information about eye velocity. Psychophysical experiments (Royden et al., 1992, 1994; Banks et al., 1996; Crowell et al., 1998) have long indicated a role for efference copy signals in heading perception. However, other psychophysical and physiological studies have demonstrated that rotational optic flow is sufficient to at least partially compensate heading estimates for eye rotation (Grigo and Lappe, 1999; Li and Warren, 2000; Sunkara et al., 2015, 2016; Manning and Britten, 2019; Burlingham and Heeger, 2020; Danz et al., 2020).

While these previous studies have demonstrated that dynamic perspective cues in optic flow are used by the brain in computations requiring estimates of eye velocity, how and where the brain computes eye rotation from optic flow has remained a mystery. Moreover, it is unknown how these visual signals about eye rotation are integrated with extraretinal (e.g., efference copy) signals regarding smooth pursuit. A good candidate substrate is the dorsal region of the medial superior temporal area (MSTd). MSTd neurons have large receptive fields (RFs; Desimone and Ungerleider, 1986; Saito et al., 1986; Komatsu and Wurtz, 1988a; Tanaka et al., 1993) and are known to respond to optic flow that simulates both translational and rotational self-motion (Saito et al., 1986; Tanaka and Saito, 1989; Duffy and Wurtz, 1991a, 1995; Gu et al., 2006; Takahashi et al., 2007). MSTd neurons are also selective for the velocity of smooth pursuit eye movements based on extraretinal signals (Newsome et al., 1988; Komatsu and Wurtz, 1988b; Ono and Mustari, 2006), suggesting the importance of the structure in tracking eye rotation. MSTd also projects back to area MT (Maunsell and Van Essen, 1983; Desimone and Ungerleider, 1986; Ungerleider and Desimone, 1986), where dynamic perspective cues are known to generate depth-sign selectivity from motion parallax (Kim et al., 2015). Together, these findings make MSTd a likely structure to process the dynamic perspective cues in rotational optic flow, and to combine visual and extraretinal signals regarding eye rotation.

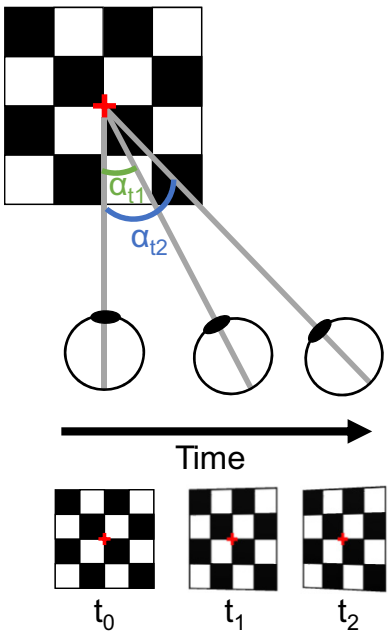


Figure 1. Schematic illustration of dynamic perspective cues. An observer translates rightward over time while maintaining fixation on a world-fixed target (red cross). As the observer translates, the eye counterrotates (angular position = α) to maintain fixation. This induces a perspective distortion of the visual scene (here a checkerboard) under planar image projection, as the gaze angle of the eye changes relative to the checkerboard. The image of the checkerboard (bottom) rotates around the fixation point such that the rightmost edge appears closer compared with the leftmost edge. This distortion is known as a dynamic perspective cue and provides information about eye rotation relative to the scene.

The main goals of this study are to determine whether MSTd neurons are tuned for the direction of smooth eye rotation, as simulated by optic flow, to compare directional tuning for eye rotation based on visual and extraretinal signals, and to quantify how MSTd neurons combine visual and nonvisual signals regarding eye rotation. Our findings provide the first evidence for a source of dynamic perspective signals that accompany eye rotation, and establish area MSTd as a key node for integrating visual and extraretinal signals regarding eye rotation that may be used in a variety of computations.

Materials and Methods

Subjects and surgery

We studied two adult male rhesus macaques (m31 and m39, *Macaca Mulatta*; weight, 10–13 kg). Standard aseptic surgical procedures under gas anesthesia were performed to implant a head restraint device. A Delrin (DuPont) ring was attached to the skull with dental acrylic cement, which was anchored by bone screws and titanium inverted T-bolts. To monitor eye movements, a scleral search coil was implanted under the conjunctiva of one eye. After recovery, subjects were trained to fixate and pursue a target for fluid rewards.

To target microelectrodes and linear array probes to area MSTd, a recording grid made of Delrin was affixed inside the head-restraint ring using dental acrylic. The grid ($\sim 2 \times 4 \times 0.5$ cm) contains a dense array of holes (spaced 0.8 mm apart). Small burr holes were drilled vertically through the recording grid to allow the entry of electrodes into the brain via sterile transdural guide tubes. The University Committee on Animal Resources approved all surgical procedures and experimental protocols at the University of Rochester.

Experimental apparatus and visual stimuli

Subjects were seated and head-fixed in a custom-made primate chair, which was then attached to a motion platform with six degrees of freedom (catalog #6DOF2000E, MOOG). A field coil frame (C-N-C Engineering) was mounted to the top of the motion platform and was used to monitor eye movements using the scleral search coil technique. For additional details of the motion platform apparatus, see Gu et al. (2006).

Visual stimuli were rear projected onto a 60 × 60 cm tangent screen using a stereoscopic projector (model Mirage S + 3K, Christie) mounted on the platform. The tangent screen was mounted on the front side of the field coil frame. To restrict the field of view of the animal to visual stimuli presented on the tangent screen, the sides and top of the field coil frame were covered with black matte material. The screen was 32 cm from the eyes of the animal and subtended ~90° × 90° of visual angle. Because of backlighting from the projector, the background luminance of the display was ~1.5 cd/m^2.

To accurately simulate the movement of the observer through a virtual environment, visual stimuli were generated using software custom written in Visual C++, using the OpenGL 3D graphics rendering library. Stimuli were rendered using a hardware-accelerated graphics card (Quadro FX 4800, NVIDIA). For conditions involving simulated eye rotation, the OpenGL camera was moved along the same trajectory of movement as the animal's eye in the real eye movement conditions, thus effectively simulating the visual consequences of eye movement (described in further detail below).

The background (when visible) was composed of a 3D cloud of "random dots," where each dot was a randomly placed triangle with a fixed base and a height of 0.22 cm. The 3D cloud was 100 cm wide, 100 cm tall, 40 cm deep, and had a density of 0.01 triangles/cm^3. Background triangles were always visible within the depth range from 12 to 52 cm from the observer. All visual stimuli, except for the fixation point, were presented monocularly to the animal as green elements against a blank background. Since the background stimulus was presented monocularly, the 3D percept of rotation around the fixation point was predominantly generated by motion parallax cues present in the stimulus. Stimuli were viewed through custom-made goggles containing red and green filters (Wratten 2 no. 29 and no. 61, Kodak).

Stimulus conditions

The main experimental protocol included the following four conditions that were randomly interleaved: Eye Only (EO), Dynamic Perspective (DP), Congruent, and Incongruent. Any real or visually simulated movement (translation or rotation) in each condition had a duration of 2000 ms and followed one cycle of a 0.5 Hz modified sinusoid. The sinusoid was modified by multiplying it by a high-power Gaussian function to smooth out the beginning and end of the movement (Nadler et al., 2009). Movements were along one of four axes, with two possible starting phases. This led to a total of eight rotation directions (Fig. 2*A*), which were common across all stimulus conditions. For example, movement along the horizontal (0°) axis with a 0° phase gives a rotation direction of 0° (Fig. 2*A*). In each trial, the fixation point first appeared directly in front of the monkey. If the condition required the monkey to track the target, then the target would move once the animal fixated. For a phase of 0°, the target would first move to the left (on the screen), then from left to right across the screen, and then from right back to the center. Data were analyzed during the middle section of the trial, 1000 ms in duration, in which the eye moved in one direction (Fig. 2*B*, between the dashed vertical lines).

Because background triangles had a fixed physical size in the virtual environment, near triangles were larger and generally moved faster (in the image on the display), while far triangles were smaller and moved more slowly. On average, the nearest triangles in the center of the display had a base and height of 1.05° and an average speed of 15.7°/s over the stimulus duration. The farthest triangles in the center of the display had a base and height of 0.24° and an average speed of 3.7°/s.

Additional trials without real or simulated rotation cues (i.e., fixation on an otherwise blank background) were interleaved to measure the spontaneous activity of each unit. After successfully completing each trial, the monkey was rewarded with a liquid reward. If the animal left the fixation window at any time during a trial, the trial was aborted and data were discarded.

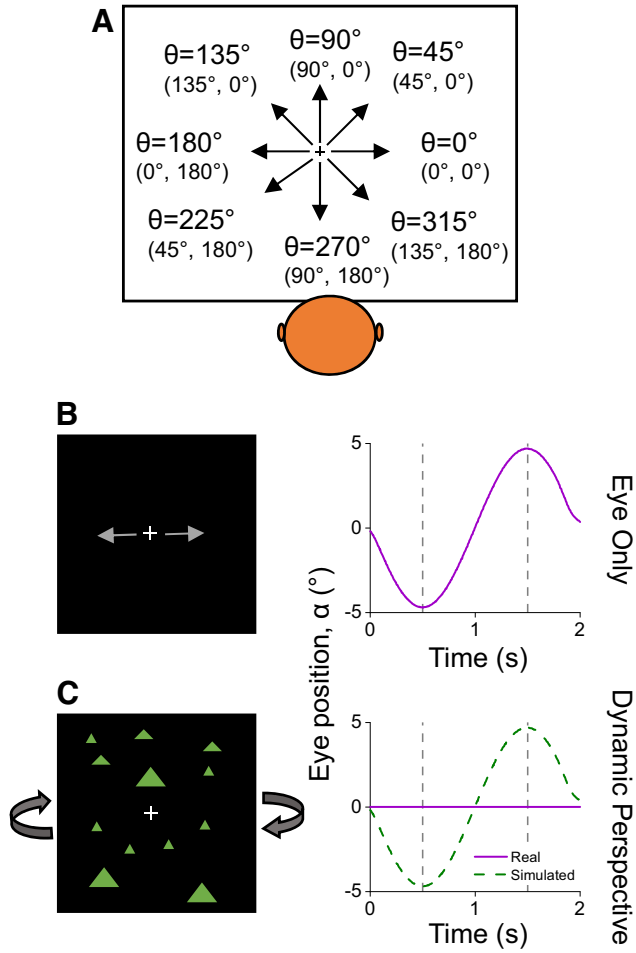
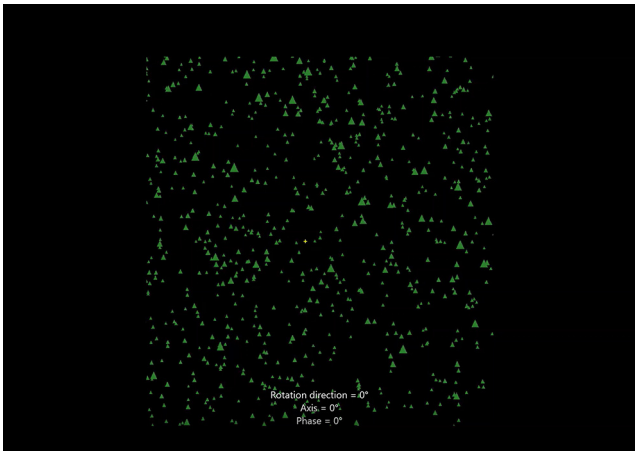


Figure 2. Definition of stimulus directions and schematic illustrations of the EO and DP stimulus conditions. ***A***, The eight rotation directions used in all stimulus conditions. Each rotation direction, θ, arises from one particular combination of 4 rotation axes and 2 rotation phases (values given in parentheses). Each trial involves one cycle of sinusoidal movement (at 0.5 Hz), and rotation direction is defined as the direction during the middle half-cycle of movement. ***B***, Eye Only condition. Left, In this illustration, the fixation target (white cross) translates along an axis of 0° and the monkey pursues the target. Background triangles are turned off. Right, Eye position as a function of time, illustrating the real (purple) eye movement. The purple curve assumes perfect pursuit such that eye position is equal to target position. As visual cues to rotation are lacking in this condition, no curve is shown corresponding to simulated eye rotation. ***C***, Dynamic Perspective condition. Left, The eye of the monkey remains fixated on a screen-centered target, while background motion simulates the combination of eye rotation and translation. Right, Visually simulated rotation (green dashed curve) follows the same sinusoidal profile as real eye rotation in ***B***, while real eye rotation is absent (purple line). Dashed vertical lines in ***B*** and ***C*** indicate the time period during which data are analyzed.

EO condition. The EO condition is intended to elicit smooth pursuit eye movements along eight directions lying within the frontoparallel plane (Fig. 2*A*), in the absence of optic flow cues to rotation. Thus, for this condition, the visible scene only consisted of a fixation point that translated laterally in front of the animal (Movie 1). The monkey was required to execute smooth pursuit eye movements to track the sinusoidal movement of the fixation target. Note that this manipulation is visually equivalent to translating the observer laterally and having them maintain fixation on a world-fixed target. There were no background triangles presented during this condition. Therefore, information about eye rotation in this condition comes from extraretinal signals associated with smooth pursuit eye movements (Fig. 2*B*). This stimulus condition

Movie 1. EO condition. In this condition, the background triangles are turned off as the fixation target (+) translates in front of the animal. The animal remains stationary and is required to make smooth pursuit eye movements to track the cross. All four rotation axes and both motion phases are shown, for a total of eight unique directions of eye rotation (as defined in Fig. 2*A*). [View online]

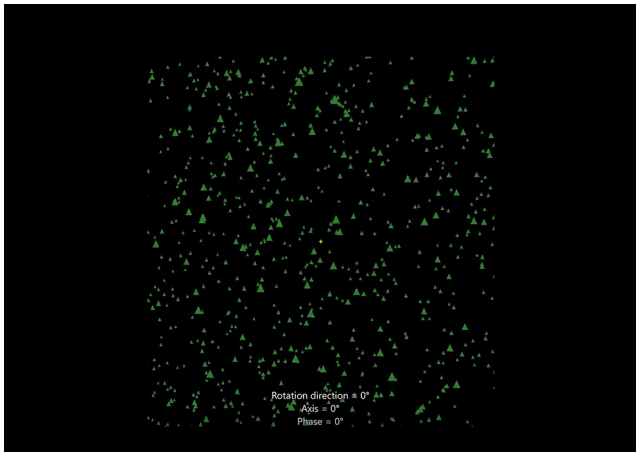


Movie 2. DP condition. In the DP condition, the background optic flow simulates the visual input received during smooth eye movements in the Congruent condition. The stationary monkey fixates on the cross in the center of the screen and passively fixates while viewing the stimulus. [View online]

is identical to the Eye Only condition used in a previous study of area MT (Nadler et al., 2009); hence, we have adopted the same nomenclature here.

Congruent condition. The Congruent condition is the same as the EO condition, except with background triangles visible. The entire 3D background, along with the fixation point, translates sinusoidally in front of the animal along an axis in the frontoparallel plane (see Fig. 7*A*, see Movie 3). In this condition, there is information about eye rotation from both extraretinal signals related to real eye rotation, as well as rotational optic flow (also known as dynamic perspective cues) generated by the visual background (see Fig. 7*C*,*E*). Note that all background triangles move in the same direction on the display at any given time; however, relative to the fixation target, near and far background triangles move in opposite directions on the retina. This condition allows us to examine how neurons in area MSTd combine congruent visual and extraretinal cues to eye rotation.

DP condition. The purpose of the DP condition is to provide optic flow (dynamic perspective) cues to eye rotation, in the absence of extraretinal signals regarding pursuit. Critically, there is no real eye rotation in the DP condition; the monkey is simply required to fixate on a target located at the center of the screen (Fig. 2*C*). To generate background



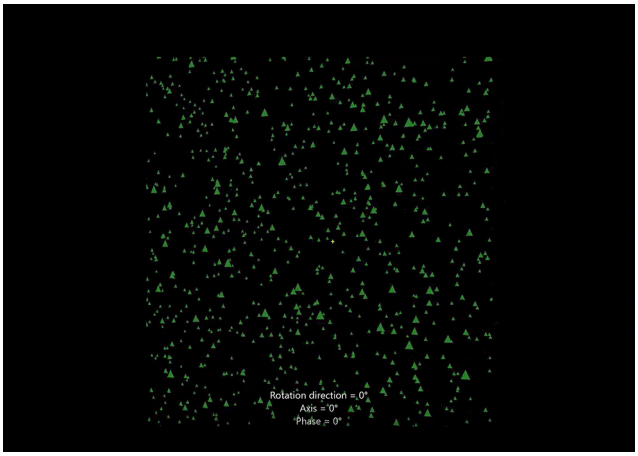
Movie 3. Congruent condition. In the Congruent condition, the background triangles and fixation point translate together along one of the four axes of motion. The stationary monkey must make smooth pursuit movements to track the fixation cross. Rotational optic flow (dynamic perspective cues) is produced by the combination of scene translation and eye rotation. [View online]

image motion that simulates optic flow induced by the same eye movements made in the Congruent condition, the OpenGL camera was translated and counterrotated (following the same 0.5 Hz sinusoidal movement profile) such that the camera was always aimed at a world-fixed point in the scene. This generates an optic flow field in which background scene elements effectively rotate around the fixation point (Movie 2). The DP condition presents the same pattern of image motion that the eye would see in the Congruent condition, assuming accurate pursuit. Previous work has shown that this rotational flow field can be interpreted by the brain as resulting from eye rotation (Kim et al., 2015; Sunkara et al., 2015). This condition is identical to the Dynamic Perspective condition used in a previous study (Kim et al., 2015); hence, the same nomenclature is adopted here.

Incongruent condition. The purpose of the Incongruent condition is to create a situation in which visual motion cues to eye rotation signal a rotation direction that is opposite to the actual direction of smooth pursuit. The Incongruent condition was identical to the Congruent condition in terms of translation of the scene relative to the animal and eye movement requirements for tracking the fixation point. However, in the Incongruent condition, the motion of the background triangles on the display screen does not simply simulate translation; rather, the background triangles rotate around the FP in the direction opposite to that produced by the real eye rotation (see Movie 4). This rotation of the background triangles is double that of the rotation of the real eye, such that the flow field produced is 180° out of phase with the real eye movement (see Fig. 7*B*). For example, when the eye translates leftward and rotates to the right in the Congruent condition (0°; see Fig. 7*C*), this produces an angle of the eye relative to the background (β_0°) that is equivalent to the angle produced in the Incongruent condition when the eye moves in the opposite direction (180°; see Fig. 7*F*). Similarly, when the eye translates rightward and rotates to the left in the Congruent condition (180°; see Fig. 7*E*), the angle of the eye relative to the background (β_{180}°) is equivalent to the angle produced in the Incongruent condition for rightward eye rotation (0°; see Fig. 7*D*). Thus, the Incongruent condition provides misaligned extraretinal and visual signals about eye rotation, allowing us to examine how the two signals may combine in the responses of MSTd neurons.

Electrophysiological recording

We recorded extracellular single unit (SU) and multiunit (MU) activity using either single tungsten microelectrodes (tip diameter, 3 μm; impedance, 0.5–2 MΩ at 1 kHz; FHC) or linear electrode arrays (24 and 32



Movie 4. Incongruent condition. Scene translation and eye movements are the same as in the Congruent condition. The key difference is that background triangles rotate around the fixation target to produce a rotational optic flow pattern that is consistent with the eye moving in the opposite direction. [View online]

channel V-probes, Plexon). Linear probes had a vertical interelectrode distance of 50 μm.

For single-electrode recordings, the sterilized microelectrode was loaded into a transdural guide tube and was advanced into the brain using a hydraulic micromanipulator (Narishige). The voltage signal of the microelectrode was amplified and filtered (300–3000 Hz; Krohn-Hite). Neural spikes were detected using a dual-window discriminator (BAK Electronics), whose output was time stamped with 1 ms resolution.

For linear array recordings, the probe was loaded into a transdural guide tube and advanced into the brain using the FlexMT microdriving terminal system (Alpha Omega). Neural signals were amplified and filtered (350–3446 Hz; Blackrock Microsystems). Activity was monitored online during the experiment, and spike sorting was subsequently performed offline (see subsection Spike sorting for linear array recordings).

Eye position signals were measured using a magnetic eye coil system (CNC Engineering) and were digitized at a sampling rate of 200 Hz for single-electrode recordings or 500 Hz for linear array recordings. Eye position signals were calibrated at the beginning of each recording session. The raw voltage signal from single microelectrodes was digitized and sampled at 25 kHz (Power1401 Data Acquisition System, Cambridge Electronic Design), whereas raw signals from linear arrays were digitized and sampled at 30 kHz (Blackrock Microsystems). For monkey m39, all units were recorded with 32-channel linear arrays across six recording sessions. For monkey m31, 35 SUs were recorded with single microelectrodes, while the remaining units were recorded with 24-channel linear arrays across eight recording sessions.

The location of area MSTd was initially identified by registering the structural MRI for each monkey with a standard macaque atlas, using CARET software (Van Essen et al., 2001). The approximate coordinates for vertical electrode penetrations were estimated from the MRI-based areal parcellation scheme, as mapped onto the MRI volume for each animal. The approximate location of area MSTd was projected onto the horizontal plane of the recording grid, and the corresponding grid holes were explored. Patterns of gray matter and white matter along electrode penetrations aided our identification of area MSTd. Upon reaching the superior temporal sulcus, we typically encountered neurons with very large receptive fields and selectivity for visual motion, as expected for area MSTd. We mapped the RFs of the MSTd neurons manually by moving a patch of drifting random dots around the visual field and observing a qualitative map of instantaneous firing rates on a custom graphical interface. MSTd neurons typically had large RFs and preferred fast speeds (>20°/s). In most cases, RFs were centered in the contralateral visual field but also extended into the ipsilateral field and included the fovea. Many of the RFs were well contained within the boundaries of our display screen, but some RFs clearly extended beyond the boundaries of the screen. Moreover, MSTd neurons usually were activated only by large visual stimuli (random-dot patch diameter, >10°), with smaller patches typically evoking little response. These properties are typical of neurons in area MSTd and are distinct from the lateral subdivision of area MST (Komatsu and Wurtz, 1988a; Tanaka et al., 1993).

To further aid identification of recording locations, electrodes were often further advanced into the middle temporal area (area MT). There was usually a quiet region 0.5–1 mm long before MT was reached, which helped to confirm the localization of MSTd. MT neurons were identified according to several properties, including smaller receptive fields (diameter approximately equal to eccentricity), sensitivity to both small and large stimuli, and gradual changes in direction preferences within electrode penetrations (Albright et al., 1984). The changes in receptive field location of MT neurons across guide tube locations were as expected from the known topography of MT (Zeki, 1974; Gattass and Gross, 1981; Van Essen et al., 1981; Maunsell and Van Essen, 1983; Desimone and Ungerleider, 1986; Albright and Desimone, 1987). Thus, we took advantage of the retinotopic organization of MT receptive fields to help confirm the locations of our electrodes within MSTd.

During most recording sessions, we first hand mapped the receptive fields and tuning properties of recorded units. This was followed by a series of quantitative measures of standard tuning properties, and then the main experimental protocol. Initially, we explored the receptive field and tuning properties using a receptive-field mapping program. This was done for all neurons recorded using single microelectrodes and for a subset of the units (typically up to four) recorded simultaneously using linear electrode arrays. A patch of moving or flickering dots was moved around the screen using a mouse, and instantaneous firing rate was plotted on a graphical user interface. This allowed us to obtain a qualitative mapping of the receptive field. Next, we estimated the preferred velocity of motion of the neuron by searching through a polar representation of direction and speed. Finally, we adjusted the horizontal disparity of an optimized patch of moving dots to assess preference for depth.

After initial hand mapping, we performed a series of quantitative tests of standard tuning properties. 2D direction tuning was measured by presenting a random-dot stimulus that drifted in one of eight different directions, 45° apart (DeAngelis and Uka, 2003). The size and location of the random-dot patch was determined from the initial receptive-field mapping. Speed tuning was measured (at the approximate preferred direction) by presenting dot patterns that drifted at 0, 0.5, 1, 2, 4, 8, 16, and 32°/s, with the same location and size used to measure direction tuning. Heading tuning from optic flow was measured by visually simulating translation along 26 heading directions corresponding to all combinations of azimuth and elevation angles in increments of 45° (Gu et al., 2006). No vestibular cues were provided for the heading tuning protocol in this study.

Next, we used a reverse-correlation technique to measure the spatial and directional receptive field structure of MSTd neurons (Chen et al., 2008). The display screen was divided into a 6×6 grid of subfields. Within each subfield, we presented a coherently moving random-dot stimulus, which could drift in one of eight directions (45° apart) on the screen. The speed of motion was fixed at 40°/s, a value that activates most MSTd neurons (Duffy and Wurtz, 1995; Churchland and Lisberger, 2005). Motion occurred simultaneously in each subfield, and the direction of motion for each patch changed randomly every 100 ms (six video frames). The direction of motion in each subfield was chosen randomly from a uniform distribution across the eight possible directions, and each subfield was updated independently of the others. Each 2 s trial thus contained a temporal sequence of 20 directions of motion within each subfield, and a new random sequence of directions was presented for each trial (Chen et al., 2008). Spatial receptive field maps obtained via reverse correlation were fit with 2D Gaussian functions to obtain receptive field sizes (full-width at half-height, averaged between the major and minor axes). The mean (±SD) receptive field size was 59 ± 17°, with a size range from 43° to 85°.

The main experimental protocol included four conditions—EO, DP, Congruent, and Incongruent—as described above. All four conditions

were randomly interleaved in a single block of trials. Spikes occurring within the middle tracking period (Fig. 2*B*,*C*, dashed vertical lines) were used in analysis. Fixation was enforced with a 3° × 3° square window for the duration of each trial. During the onset of tracking in conditions that involved real pursuit, the fixation window size was initially 4° × 4° to allow an initial catch-up saccade (if necessary). After 250 ms, the window shrunk to 3° × 3° for the remainder of the trial duration. Any deviation of the eye position from the window led to the trial being aborted and data being discarded. All MU and SU activity in MSTd were analyzed after offline spike sorting.

Spike sorting for linear array recordings

All data collected with linear electrode arrays went through an automatic spike-sorting procedure, followed by manual curation using Offline Sorter (Plexon). Each channel yielded one MU with a possibility of one or more SUs. For each channel, candidate spikes were detected when the raw voltage trace was 3 SDs below the mean voltage. The waveforms of the detected candidate spikes were then sorted into units using the *t*-distribution expectation–maximization scanning method, a built-in feature of the Offline Sorter software, and principal component analysis (PCA) was performed for sorted units. Units were joined together as one MU if not clearly separated in the PCA-based feature space. Cross-correlograms were computed across channels for each recording to ensure that a SU was not counted on multiple channels. Spikes were also examined across tuning protocols to ensure the spike stayed stable across time. This served to ensure the validity of using tuning parameters across protocols in the same recording session.

Data analyses

Analyses of spike data were performed using custom software written in MATLAB (MathWorks). Statistical tests were performed in MATLAB and OriginPro (OriginLab). Basic neural tuning properties were quantified and analyzed. Tuning curves for 2D direction of motion were calculated for each unit (mean firing rates as a function of motion direction) and fit to wrapped Gaussian functions (Fetsch et al., 2007). Speed tuning curves were calculated for each unit (mean firing rates as a function of speed) and fit to a Gamma function (DeAngelis and Uka, 2003; Nover et al., 2005).

For the main experimental protocol, mean firing rates (spikes per second) as a function of rotation direction were calculated for each stimulus condition to quantify the tuning properties of units for visually simulated and real eye rotations. The time window for calculating the mean firing rate began 150 ms after the start of the middle tracking period (Fig. 2*B*,*C*, vertical dashed lines) and ended 150 ms after completion of the middle tracking period (between 650 and 1650 ms). This 150 ms delay relative to the middle tracking period served to compensate for the latency of pursuit eye movements (Robinson, 1965; Krauzlis and Miles, 1996). This latency was chosen to avoid contamination of firing rates from the opposite phase of eye rotation.

Tuning curves for rotation direction (mean firing rates as a function of rotation direction) in the EO and DP conditions were fit to a von Mises function of the following form:

$$R(\theta) = R_0 + A * e^{\left(-2*\frac{1-\cos(\theta-\theta_0)}{\sigma^2}\right)}, \tag{1}$$

where R_0 denotes baseline response level, A represents tuning curve amplitude, θ_0 denotes the preferred direction (in degrees), and σ represents tuning width. Tuning curves were additionally fit to a bimodal von Mises function of the following form:

$$R(\theta) = R_0 + A * e^{\left(-2*\frac{1-\cos(\theta-\theta_0)}{\sigma^2}\right)} + A_2 * e^{\left(-2*\frac{1-\cos(\theta-\theta_0-180)}{\sigma_2^2}\right)}, \tag{2}$$

where A_2 and σ_2 correspond to the amplitude and tuning width, respectively, of the second peak. A log-likelihood ratio test (χ^2 test, $p < 0.05$) was performed to determine which model best fit the data for each unit. Classification of tuning curves as bimodal also required that the amplitude of the secondary peak was at least 20% of the amplitude of the primary peak.

To quantify the ability of each unit to discriminate between its preferred and nonpreferred rotation directions, a direction discrimination index (DirDI) was computed (Prince et al., 2002; DeAngelis and Uka, 2003; Takahashi et al., 2007; Danz et al., 2020), as follows:

$$\mathrm{DirDI} = \frac{\mathrm{Rmax} - \mathrm{Rmin}}{\mathrm{Rmax–Rmin} + 2\sqrt{\left(\frac{\mathrm{SSE}}{(N-M)}\right)}}, \tag{3}$$

Where Rmax and Rmin indicate the maximum and minimum responses, SSE denotes the sum of squared error, N is the number of trials, and M is the number of rotation directions. This index ranges from 0 to 1 (weak to strong discrimination ability) and quantifies the strength of directionality relative to the intrinsic response variability of the unit.

Linear summation model

The EO and DP conditions provided isolated signals of real or visually simulated eye rotation, allowing us to measure neural responses to each source separately. To better understand how MSTd neurons combine these visual and extraretinal signals about eye rotation, we used a linear model to predict neural responses to the Congruent and Incongruent conditions from measured responses in the EO and DP conditions. Units with significant tuning (ANOVA, $p < 0.05$) in the EO and DP conditions, as well as in the direction and speed tuning measurements, were included in this analysis. The linear model followed the form:

$$\mathrm{R_{pred}}(\theta) = (\mathrm{R_{EO}}(\theta) * \mathrm{W_{EO}}) + (\mathrm{R_{DP}}(\theta) * \mathrm{W_{DP}}) + (\mathrm{R_{RS}}(\theta) * \mathrm{W_{RS}}), \tag{4}$$

where $\mathrm{R_{pred}}(\theta)$ represents the predicted responses for each rotation direction (θ) in the Congruent and Incongruent conditions; R_x denotes measured response (mean spike rate) from the EO condition, the DP condition, or the retinal slip (RS) predictor (described below); W_x denotes the fitted weight for each of these predictors. A single set of weights was obtained to predict responses for both the Congruent and Incongruent conditions (by minimizing errors pooled across those conditions), since we presume that each neuron has a fixed combination rule across stimulus conditions. For the Incongruent condition, the response predictor from the DP condition, $\mathrm{R_{DP}}$, was shifted by 180° (see Figs. 8*B*, 9*B*, gray lines). This 180° shift accounts for the stimulus manipulation of the simulated rotation cues present in the Incongruent condition. We used the MATLAB optimization function *fmincon* to find weights that minimize SSE between the predicted and actual responses in the Congruent and Incongruent conditions.

Model evaluation

To evaluate the fit of our linear model for each MSTd neuron, we calculated the variance accounted for (VAF) by the model, with and without the retinal slip predictor, using the ratio of SSE to the sum of squared total (SST) response. SSE quantifies the error between the data and model predictions, as follows:

$$\mathrm{SSE} = \sum_{i=1}^{M} (y_i - \hat{y})^2, \tag{5}$$

where M denotes the number of rotation directions, y represents the measured neural response, and $\hat{y}$ denotes the predicted response. SST quantifies the variation in the measured responses as the squared difference between the observed response for each rotation direction and the mean response across rotation directions, $\bar{y}$:

$$\mathrm{SST} = \sum_{i=1}^{M} (y_i - \bar{y})^2. \tag{6}$$

The ratio of SSE to SST quantifies the badness of fit of the model, with low values being better. VAF was computed as follows:

$$\text{VAF} = 1 - \frac{SSE}{SST}. \quad (7)$$

Higher values of VAF indicate that the predictions for the Congruent and Incongruent conditions are better matched to the actual responses in those conditions.

Computation of the retinal slip predictor

In principle, responses to the Congruent and Incongruent conditions should be predictable by some function of responses to the EO and DP conditions. However, this effectively assumes that smooth pursuit is perfect. Given that pursuit in the EO, Congruent, and Incongruent conditions cannot be perfect (there must at least be some lag), then there may also be a substantive contribution to MSTd responses caused by retinal slip of the pursuit target. To attempt to account for this effect, a RS predictor was included in our linear model, as described above. The significance of the contribution of the RS predictor was assessed using a sequential F test ($p < 0.05$).

To formulate a RS predictor, we first computed the average retinal image slip (across trials) of the fixation target in the Congruent and Incongruent conditions, separately for each rotation direction. Retinal slip was calculated as the difference between target and eye velocity signals. For this purpose, vertical and horizontal eye position signals were linearly interpolated to a resolution of 1 ms, filtered with a Gaussian window having $\sigma = 33$ ms (MATLAB function *filter*) and then differentiated to obtain eye velocity. For each trial, the average difference between target velocity and eye velocity was calculated during the middle tracking period (same time window over which firing rates were computed). This retinal slip vector was then converted to a direction and speed of slip, and the expected neural response to this retinal slip was computed from the direction and speed tuning curves that were measured independently for each neuron. By assuming that direction and speed tuning are separable (Rodman and Albright, 1987), we could then predict the response to any retinal slip from the product of the fitted direction and speed tuning curves. The average predicted response for each rotation direction constituted the RS predictor in Equation 4 (see also Fig. 9*C*,*D*). Note that the overall scale of the RS predictor does not matter much as it is multiplied by a fitted weight; thus, the shape of the RS predictor is the key element.

For a handful of neurons, the standard 2D direction tuning protocol was not run, but a 3D visual heading tuning condition was performed (Gu et al., 2006). In these cases, direction tuning was obtained by taking the responses to heading directions within the frontoparallel plane. For neurons with both standard 2D direction tuning and 3D heading tuning, there was a strong correlation between direction preferences obtained from the two protocols ($\rho = 0.877$, $p = 1.62 \times 10^{-11}$, circular–circular correlation).

Results

We measured responses of MSTd neurons to real eye rotations and to stimuli that visually simulate the optic flow produced by eye rotations, as well as to congruent and incongruent combinations of the two. We first describe how single neurons respond to these real and visually simulated eye rotations in isolation and compare the rotation tuning of neurons in these two conditions. We next examine how neurons respond to combinations of real and simulated eye rotations that are either congruent or incongruent. Finally, we investigate whether neural responses to these combinations can be predicted simply from responses to the real and simulated rotations presented in isolation.

Two monkeys were trained to maintain visual fixation on a target that could be either stationary or smoothly moving relative to the head of the animal. Target movement followed a modified 0.5 Hz sinusoid (see Materials and Methods), and this produced sinusoidal eye rotation along one of four axes. Two possible starting phases of motion were used, giving a total of eight directions of eye rotation (Fig. 2*A*). In some task conditions, a combination of eye translation and rotation was visually simulated (following the same 0.5 Hz sinusoid), while the animal maintained fixation on a screen-fixed target. Neural responses were measured during the middle section of the trial, when the eye was rotating in a consistent direction across the screen (Fig. 2*B*,*C*, vertical dashed lines).

All task conditions involved a viewing geometry in which the eye translated relative to the scene while making a compensatory smooth pursuit movement to maintain gaze on a world-fixed point (Fig. 1). Note that the head and body remained stationary in these experiments, such that translation of the eye was visually simulated in all conditions. However, the task conditions varied in terms of the axis of eye rotation and whether the eye rotations were real or visually simulated. When a visual background scene was present, it consisted of a 3D cloud of triangles (for details, see Materials and Methods). In the EO condition, there was no background scene other than the fixation point, which translated sinusoidally along an axis within the frontoparallel plane of the display (for details, see Materials and Methods; Movie 1). The animal was required to make smooth pursuit eye movements to track the fixation point. Since the background triangles were off in this condition, there were no visual cues to eye rotation (i.e., dynamic perspective cues) from background motion (Fig. 2*B*). In the DP condition, the stimulus provided visual cues to eye rotation without real eye rotation. In this case, the visual scene again translated relative to the eye, and we visually simulated (by rotating the OpenGL camera) a compensatory counterrotation of the eye, such that gaze remained fixed on a world-fixed target (see Materials and Methods). This generated an optic flow pattern in which background elements effectively rotated around the point of fixation (Fig. 2*C*, Movie 2), while the eye remained stationary relative to the head (for details, see Materials and Methods). The rotation directions simulated in this condition were the same as those used in the EO condition.

We first examine MSTd responses to the EO and DP conditions, and we return later to examine responses to Congruent and Incongruent combinations of the visual and extraretinal cues to eye rotation. We spike sorted and analyzed all MSTd units collected in each recording session. This included both MU and SU activity. MU activity reflects the summed activity of nearby units on a particular channel. All SU spikes were removed from the MU activity when both occurred on the same channel, such that SU and MU signals recorded from the same channel may be considered independent (Chen et al., 2008; see also Materials and Methods).

Comparison of multiunit and single unit tuning for direction of eye rotation

We first asked how tuning properties compare between MU and SU activity to assess whether MU signals are representative of clustered SU responses to our stimuli, as shown previously for heading tuning in area MSTd (Chen et al., 2008; Shao et al., 2018). We compared MU and SU responses for the 2D motion direction and speed tuning curves, as well as for tuning to rotation direction in the EO and DP conditions.

Figure 3 shows tuning curves for a typical recording of MU (Fig. 3, left column) and SU (Fig. 3, right column) activity from the same channel during one recording session. Standard 2D (frontoparallel) direction tuning was measured in response to a patch of random dots that moved in one of eight directions, 45° apart, at a fixed speed. Figure 3*A* shows that direction tuning was very similar for MU and SU activity recorded at this site, with a preference for rightward 2D motion (0/360°). Speed tuning was

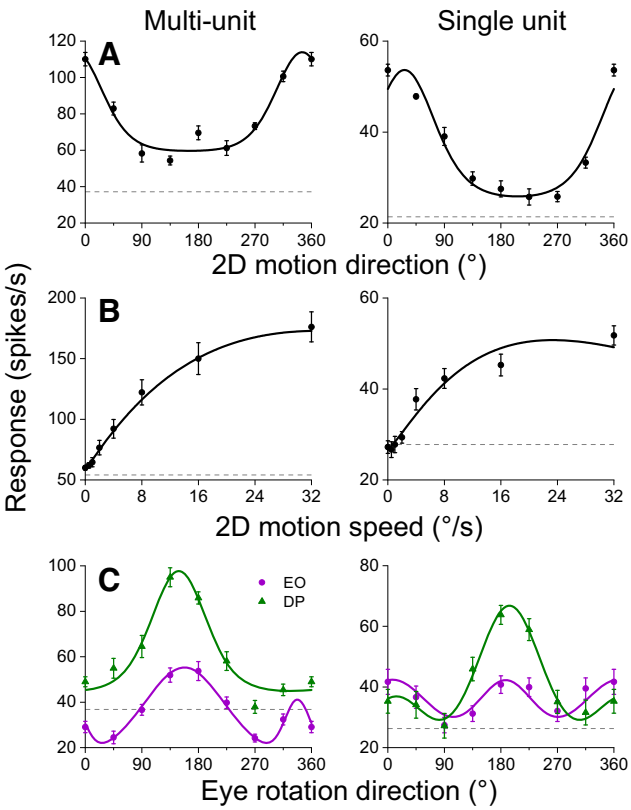


Figure 3. Data from an example MU (left column) and SU (right column) pair recorded simultaneously. ***A***, Mean spike rates are plotted as a function of 2D motion direction and fit with a wrapped Gaussian function (solid curves). ***B***, Mean spike rates are plotted as a function of 2D motion speed and fit with a Gamma function (solid curves). ***C***, Mean spike rates are plotted as a function of real and simulated rotation direction for the EO (purple circles) and DP (green triangles) conditions, respectively. Smooth curves show fits of a unimodal or bimodal von Mises function (see Materials and Methods). Error bars in all panels indicate SEM.

measured with random dots that moved in a fixed direction at one of eight speeds (0, 0.5, 1, 2, 4, 8, 16, and 32°/s). Again, MU and SU activity showed very similar tuning for speed (Fig. 3*B*), with a preference for fast image speeds, as is typical for MSTd neurons (Tanaka and Saito, 1989; Chukoskie and Movshon, 2009; Inaba et al., 2011).

To calculate rotation tuning curves for the main stimulus conditions, mean firing rates were plotted as a function of rotation direction (Fig. 2*A*, definition of rotation directions) for both the EO and DP conditions. As shown in Figure 3*C* for the example MU/SU recording, tuning for the direction of eye rotation was similar between MU and SU responses for both the EO and DP conditions, with a preference for leftward eye rotation (~180°). Thus, we found strong similarity of tuning properties between MU and SU responses from the same recording site, as further quantified below at the population level.

Neural tuning for both real and visually simulated eye rotations was common in area MSTd, in both MU and SU responses. Figure 4*A*–*D* shows the rotation direction preferences of all MU and SU recordings that exhibited significant tuning for eye rotation (ANOVA, $p < 0.05$) in the EO and DP conditions. Of a total of 561 units (440 MU, 121 SU) recorded in two animals, 297 (225 MU, 71 SU) had significant rotation tuning in the EO condition, 249 (181 MU, 67 SU) had significant rotation tuning in the DP condition, and 194 (140 MU, 54 SU) had significant tuning in both conditions. In other words, ~50% of units were selective for the direction of real or simulated eye rotations, and ~35% of units were selective for both. While selectivity for the direction of smooth pursuit is well established in area MSTd (Newsome et al., 1988; Komatsu and Wurtz, 1988b; Ono and Mustari, 2006), the substantial percentage of units exhibiting rotation tuning in the DP condition supports the notion that MSTd carries a more generalized representation of eye rotation in which neurons respond to rotational optic flow that simulates the visual consequences of combined eye translation and counterrotation. With the exception of one preliminary and qualitative report (Saito et al., 1986), this form of visual rotational selectivity has not previously been studied systematically in area MSTd (see Discussion).

Across the population, direction preferences for real and simulated eye rotation were broadly distributed, but we observed a bias toward leftward and downward rotation preferences in both EO and DP conditions. This was true for both SU (Fig. 4*A*, *B*) and MU (Fig. 4*C*,*D*) activity. Since all neurons were recorded in the left hemisphere of both monkeys, the leftward bias observed is ipsiversive. Previous work has also shown biases for rotation in the downward and ipsiversive directions in the floccular lobe of the cerebellum (Krauzlis and Lisberger, 1996), and biases for ipsiversive directions in MST (Squatrito and Maioli, 1997; Ilg and Thier, 2003).

Next, we directly compared the rotation preferences of MU and SU activity recorded simultaneously on the same channel to assess clustering of rotation tuning. Direction preferences of MU and SU activity were very similar in the EO condition, with the vast majority of data points lying within 45° of the identity line (Fig. 4*E*). The difference distribution (Fig. 4*E*, inset) displays a clear peak around a difference of 0°, and the MU and SU preferences are strongly correlated ($\rho = 0.81$, $p = 1.10 \times 10^{-5}$, circular–circular correlation). Very similar results were seen in the DP condition (Fig. 4*F*), where there was again a very robust correlation between SU and MU preferences ($\rho = 0.83$, $p = 1.23 \times 10^{-4}$, circular–circular correlation). These results strongly support clustering of neurons in area MSTd according to their direction tuning for real and visually simulated eye rotations.

Comparison of MSTd preferences for eye rotation in EO and DP conditions

We next asked the crucial question of how responses of MSTd neurons to real and simulated eye rotation compare to one another. Do MSTd neurons have similar rotation tuning for optic flow and extraretinal signals? First, we examined response strength across the EO and DP conditions by calculating the peak response minus trough response from each tuning curve (Fig. 5*A*). We fit a generalized linear regression model (MATLAB *fitglm*) to the square root of peak–trough responses (the square root was taken to improve normality), with peak–trough response in the EO condition as the dependent variable. The three independent variables were as follows: peak–trough response in the DP condition, monkey identity (m39/m31), and unit type (MU/SU). We found a significant main effect of peak–trough response in the DP condition ($t_{(554)} = 28.3$, $p = 1.14 \times 10^{-109}$), indicating that response strength is well correlated across EO and DP conditions. We also found a significant main effect of monkey ($t_{(554)} = 5.86$, $p = 8.03 \times 10^{-9}$) and a significant interaction between task condition and monkey ($t_{(554)} = -11.8$, $p = 8.65 \times 10^{-29}$). These effects stem from the clear observation that m31 shows stronger peak–trough responses in the EO condition (blue data points above the diagonal), whereas m39 shows stronger peak–trough

responses in the DP condition (Fig. 5*A*, black data points below the diagonal). No significant main effect or interactions with unit type (MU/SU) were found. Despite some animal differences, this analysis demonstrates that the strength of rotation tuning is strongly related between the EO and DP conditions (Fig. 5*A*).

Next, we compared rotation tuning strength between the EO and DP conditions (Fig. 5*B*), using a DirDI that quantifies the ability of a unit to discriminate between different rotation directions relative to its intrinsic variability (for details, see Materials and Methods; DeAngelis and Uka, 2003). We used a similar generalized linear regression model, with DirDI in the EO condition as the dependent variable, and DirDI in the DP condition, monkey (m39/m31), and unit type (MU/SU) as independent variables. We found a significant main effect of DirDI in the DP condition ($t_{(554)} = 7.00$, $p = 7.50 \times 10^{-12}$), with no main effects of monkey or unit type. There was a significant interaction between condition and monkey ($t_{(554)} = -4.23$, $p = 2.68 \times 10^{-5}$), reflecting the observation that DirDI was slightly greater in the EO condition for m31 (Fig. 5*B*, blue symbols). Overall, the significant main effect of condition reflects a robust correlation between DirDI in the EO and DP conditions. This provides further evidence that MSTd carries robust signals regarding eye rotation based on both visual and extraretinal cues.

We also compared the width of tuning between the EO and DP conditions by computing tuning bandwidth as the full width of the tuning curve at half-height. We find a significant difference between tuning widths in the EO and DP conditions ($p = 8.91 \times 10^{-4}$, Wilcoxon signed-rank test), with narrower tuning for the DP condition (median, 72.9°) than the EO condition (median, 87.1°).

Critically, to evaluate whether rotation direction preferences of MSTd units are similar for real and visually simulated eye rotations, we compared rotation direction preferences between EO and DP conditions for all units with significant tuning for both conditions ($p < 0.05$, ANOVA). Figure 5, *C* and *D*, shows that the vast majority of both MUs and SUs have very similar direction preferences for eye rotation in the EO and DP conditions. The insets in Figure 5, *C* and *D*, display the difference distributions, which are clustered around zero. Direction preferences between the EO and DP conditions are highly correlated for both MU activity ($\rho = 0.83$, $p < 10^{-324}$, circular–circular correlation) and SU activity ($\rho = 0.93$, $p = 4.53 \times 10^{-9}$, circular–circular correlation). These findings show clearly that direction preferences for real and simulated eye rotation are highly similar in MSTd.

What is the relationship between preferences for eye rotation in the EO/DP conditions and the direction tuning of a neuron for 2D image motion? Previous studies of MSTd neurons have generally found that the direction preference for 2D motion tends toward being opposite (180° difference) to the direction preference for pursuit eye movements when tracking an isolated target (Newsome et al., 1988; Komatsu and Wurtz, 1988a; Ono and Mustari, 2006; Ono et al., 2010; but also see Squatrito and Maioli, 1997). Thus, we expected direction preferences in the EO condition to be the opposite of 2D direction tuning preferences. Indeed, Figure 5*E* shows that this expectation was generally confirmed. Although there is a range of relative direction preferences, as also seen in previous studies cited above, direction preferences for the EO condition are

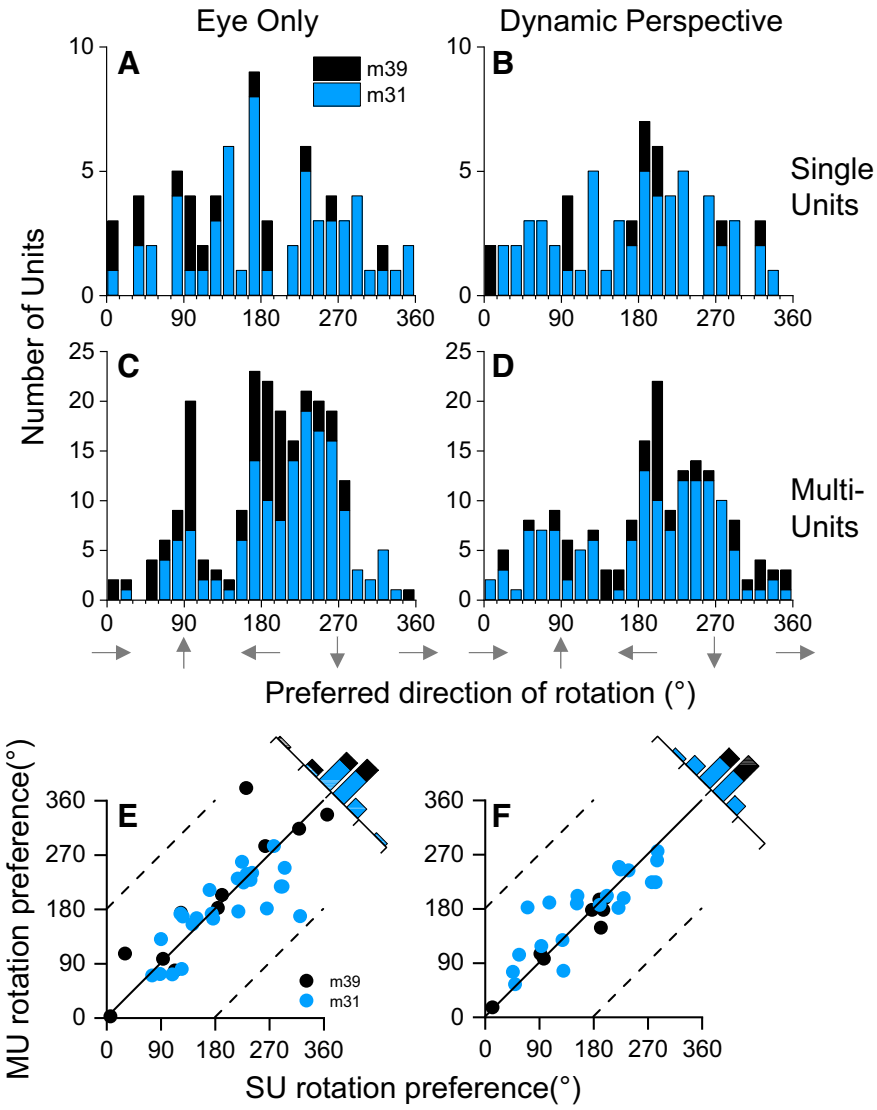


Figure 4. Summary of direction preferences for eye rotation in the EO (left column) and DP (right column) conditions. ***A***, Distribution of rotation direction preferences for SUs recorded in the EO condition. Only SUs with significant tuning for rotation direction (ANOVA, $p < 0.05$) in this condition were included. Data from monkeys m39 (black) and m31 (blue) are shown separately ($N = 16$ for m39, $N = 55$ for m31). ***B***, Distribution of SU rotation preferences for the DP condition. SUs with significant rotation tuning (ANOVA, $p < 0.05$) in the DP condition were included ($N = 12$ for m39, $N = 55$ for m31). ***C***, Same as ***A***, but for MU activity ($N = 78$ for m39, $N = 147$ for m31). ***D***, Same as ***B***, but for MU activity ($N = 48$ for m39, $N = 133$ for m31). ***E***, Comparison of rotation direction preferences MU and SU activity recorded from the same channel. Data were included if both MU and SU responses had significant rotation tuning (ANOVA, $p < 0.05$) in the EO condition ($N = 11$ for m39, $N = 25$ for m31). Data points in the scatter plot were wrapped to lie between diagonal lines corresponding to differences of ±180°. Inset displays the difference distribution between MU and SU direction preferences. ***F***, Same as ***E*** but for the DP condition. Channels were included if both MU and SU responses had significant rotation tuning (ANOVA, $p < 0.05$) in the DP condition ($N = 8$ for m39, $N = 21$ for m31).

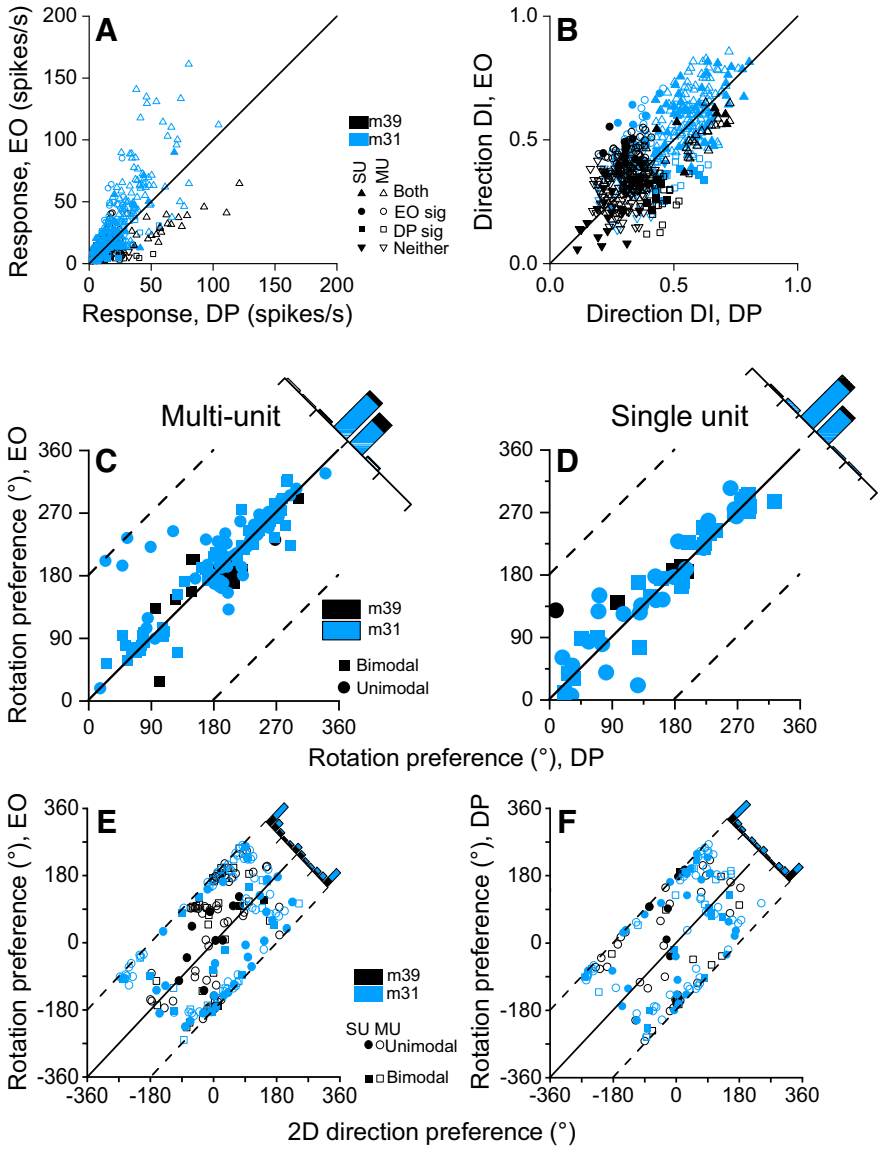


Figure 5. Comparison of rotation direction tuning properties between EO and DP conditions. ***A***, The magnitude of response of each unit (difference in response between the peak and trough of the rotation tuning curve) in the EO condition is plotted against response magnitude in the DP condition. Color denotes monkey (black for m39, blue for m31), shape denotes rotation tuning significance [upward triangle, EO and DP tuning both significant ($N = 30$ for m39, $N = 165$ for m31); circle, only EO tuning significant ($N = 65$ for m39, $N = 37$ for m31); square, only DP tuning is significant ($N = 31$ for m39, $N = 23$ for m31); downward triangle, the tuning of neither condition was significant ($N = 150$ for m39, $N = 61$ for m31)], and symbol fill denotes MU or SU (solid for SU, open for MU). Solid line indicates the unity–slope diagonal. All MSTd units were included regardless of significance of rotation tuning. ***B***, Comparison of tuning strength, as quantified by the DirDI, between EO and DP conditions. Format is the same as in ***A***. All MSTd units were included regardless of the significance of rotation tuning. ***C***, Comparison of rotation direction preferences between EO and DP conditions. This panel includes MU activity recorded from m39 (black, $N = 23$) and m31 (blue, $N = 117$). For units with bimodal direction tuning in either condition (71 of 140), the smallest difference between peaks in the two conditions was used. All MUs with significant rotation tuning (ANOVA, $p < 0.05$) in both the EO and DP conditions were included. Inset displays the difference distribution between rotation preferences for the two conditions. Format is the same as in Figure 4*E*. ***D***, Same as ***C*** but for SUs ($N = 7$ for m39, $N = 48$ for m31) with significant rotation tuning (ANOVA, $p < 0.05$) in both EO and DP conditions. Format is the same as in ***C***; 29 of 55 SUs had bimodal direction tuning. ***E***, Rotation preferences of units in the EO condition plotted against their preferences for 2D motion direction ($N = 94$ for m39, $N = 202$ for m31). Units with significant tuning in the EO condition (ANOVA, $p < 0.05$) and 2D direction tuning protocol (ANOVA, $p < 0.05$) were included (34 of 94 with bimodal EO tuning for m39; 49 of 202 bimodal for m31). Inset, Histogram (top right) shows the distribution of differences between EO and 2D direction preferences. For units with bimodal tuning in the EO condition, the direction at the largest peak was used. Color denotes monkey (black for m39, blue for m31), symbol fill denotes unit type (open for MU, solid for SU), and symbol shape denotes tuning modality (circle for unimodal, square for bimodal). Note that axis ranges are expanded such that all data points wrap between the black diagonal lines corresponding to differences of ±180°. ***F***, Same as in ***E*** but for rotation direction preference in the DP condition ($N = 60$ for m39, $N = 188$ for m31). Units with significant tuning in the DP condition (ANOVA, $p < 0.05$) and significant tuning for the 2D direction tuning protocol (ANOVA, $p < 0.05$) were included (12 of 60 with bimodal DP tuning for m39; 61 of 188 bimodal for m31). Format is the same as in ***E***.

close to 180° apart from 2D direction preferences for most neurons. The distribution of these differences was significantly different from a uniform distribution, which would indicate no relationship ($p = 1.929 \times 10^{-17}$, Kolmogorov–Smirnov test).

As expected from the close alignment of rotation preferences in the EO and DP conditions (Fig. 5*C*,*D*), we also found that DP direction preferences tend to be opposite to the direction preferences for 2D motion (Fig. 5*F*). Again, the difference distribution was significantly different from uniform ($p = 1.51 \times 10^{-18}$, Kolmogorov–Smirnov test). In the following section, we consider further the basis for this relationship.

Together, the results of Figure 5 show that responses of MSTd units in the EO and DP conditions are very similar in their response magnitude, direction discriminability, rotation direction preference, and relationship to 2D motion preference. These similarities strongly suggest a role of MSTd in representing eye rotation via both extraretinal and visual signals.

What visual cues underlie tuning for eye rotation in the Dynamic Perspective condition?

It is well established that neurons in area MSTd are tuned for the direction and speed of 2D motion (Saito et al., 1986; Komatsu and Wurtz, 1988a; Tanaka and Saito, 1989; Duffy and Wurtz, 1991a; Churchland and Lisberger, 2005; Inaba et al., 2007, 2011; Chukoskie and Movshon, 2009), as well as depth from binocular disparity (Roy et al., 1992; Takemura et al., 2000; Yang et al., 2011). In the stimulus for our DP condition, the direction and speed of motion of individual background triangles depend on their location in depth. Specifically, triangles that are nearer and farther than the fixation point move in opposite directions, and their speed increases with distance in depth away from the fixation point (Movie 2). Thus, we now consider whether the rotation tuning observed in the DP condition can be explained by basic visual response properties of MSTd neurons.

Consider the case of an MSTd neuron with a receptive field in the right hemifield, and a rightward (0°) 2D motion preference (Fig. 6*A*). Based on previous literature (Newsome et al., 1988; Komatsu and Wurtz, 1988a; Ono and Mustari, 2006; Ono et al., 2010) and Figure 5*E*, we expect such a neuron to prefer leftward

(180°) pursuit eye movements. Since a pure leftward eye rotation (without eye translation) induces image motion of background elements in the opposite direction (Fig. 6*B*), the 2D motion and pursuit preferences of such neurons can be considered congruent with respect to the retinal image motion experienced.

Now consider the simulated viewing context of this study, in which the eye translates (e.g., rightward) while counterrotating (e.g., leftward) to maintain fixation on a world-fixed point (Fig. 6*C*). We have shown in Figure 5*F* that MSTd neurons tend to prefer simulated eye rotations in the DP condition that match the rotation preference of the EO condition and are opposite to the 2D motion preference. Thus, in the illustration of Figure 6*C*, we would expect the hypothetical MSTd neuron with a rightward 2D motion preference to prefer leftward simulated eye rotation in the DP condition. In this case, while we simulate rightward eye translation and simultaneous leftward counterrotation (Fig. 6*C*), near objects that are stationary in the scene move leftward in the image, and far objects move rightward. Additionally, near objects are larger and have a substantially greater range of speeds than far objects (for details, see Materials and Methods).

Given these facts, rotation preferences in the DP condition are difficult to explain in terms of 2D motion tuning. For the hypothetical case illustrated in Figure 6*A*–*C*, a neuron that prefers rightward 2D motion would be expected (based on the data in Fig. 5*F*) to prefer a DP stimulus in which the larger and faster near triangles of the background move leftward in the image; that is, in the nonpreferred direction of the unit. The only way to explain the relationship between DP rotation preferences and 2D direction preferences (Fig. 5*F*), without invoking a mechanism beyond simple 2D motion tuning, would be that MSTd responses are driven predominantly by the image motion of far objects in the 3D cloud of triangles, since these far triangles would move in the preferred 2D motion direction of the neuron (Fig. 2*C*). Given that the far triangles are smaller and move more slowly in the image (see Materials and Methods), and that MSTd neurons tend to prefer faster speeds (Tanaka and Saito, 1989; Chukoskie and Movshon, 2009; Inaba et al., 2011), this is a very unlikely explanation. Furthermore, given that the 3D cloud of triangles was presented monocularly in our experiments, it is not possible that MSTd neurons could be selectively responding to the far triangles because of a far preference for binocular disparity.

To more quantitatively assess whether DP tuning could be explained by 2D direction and speed tuning, we computed the absolute value of the angular difference between the rotation preference in the DP condition and the 2D motion direction preference, and plotted this difference as a function of speed preference (Fig. 6*D*). If MSTd responses are simply explained by 2D image motion, then neurons with large absolute angular differences should prefer slow speeds such that they respond preferentially to the far background triangles, and neurons with small angular differences should prefer faster speeds such that they respond best to near triangles (i.e., a negative correlation between angular difference and speed preference). Inconsistent with this explanation, we found that most MSTd neurons in our sample prefer fast speeds, which is consistent with the literature (Tanaka and Saito, 1989; Chukoskie and Movshon, 2009; Inaba et al., 2011), and have nearly opposite direction preferences for rotation in the DP condition and 2D motion (Fig. 6*D*). Critically, we find no correlation between absolute angular differences and preferred speeds across the population ($p = 0.318$, pooled across animals; $p = 0.34$ for m39; $p = 0.19$ for m31; Spearman's rank correlations).

To further assess whether it is plausible that DP responses could be selectively driven by the 2D image motion of far triangles in the background, we estimated the expected relative responses of MSTd neurons to near and far triangles based on speed tuning curves. As a reasonable first-order approximation of the neural responses to near and far triangles in the 3D cloud, we integrated the area under each speed tuning curve up to the maximum speed of the far triangles (3.7°/s), and up to the maximum speed of the near triangles (15.7°/s). We find that the expected response to near triangles far exceeds the expected response to far triangles for all neurons (Fig. 6*E*). Data points

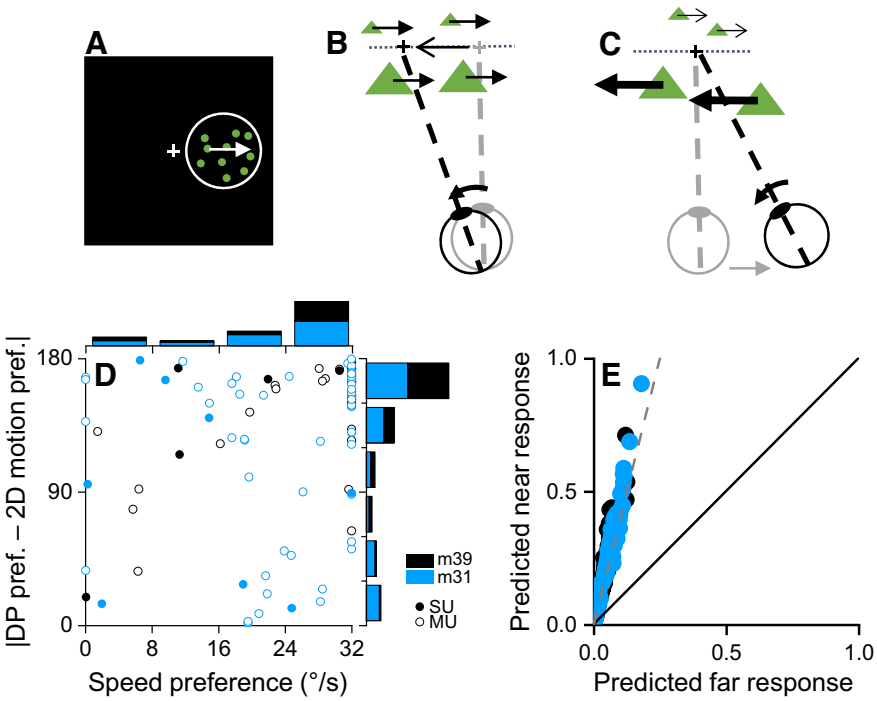


Figure 6. Are responses in the DP condition driven by selectivity for 2D motion? ***A***, Schematic depiction of a stimulus for measuring tuning for 2D motion direction. The white circle indicates the receptive field, and the white arrow indicates the preferred 2D motion direction (rightward, 0°). ***B***, A pure leftward rotation of the eye produces rightward image motion of the stationary background elements (green triangles). ***C***, When the eye translates rightward while counterrotating to the left to maintain fixation on a world-fixed target (black cross), near and far background elements in the scene (green triangles) move in opposite directions and with speeds that depend on their distance relative to the fixation point. ***D***, The absolute difference between rotation direction preferences in the DP condition and 2D motion direction preferences (|DP rotation − 2D motion preference|) is plotted as a function of speed preference for a subpopulation of MSTd neurons. Marginal histograms show distributions of direction differences and speed preferences. All MSTd units (solid fill for SU, open fill for MU) from m39 (black) and m31 (blue) with significant tuning for speed and for rotation in the DP condition (ANOVA, $p < 0.05$) were included. ***E***, Expected responses for near triangles are plotted against expected responses for far triangles (see text for details). Same subpopulation of units as in ***D***. The black line denotes the unity slope diagonal, while the dashed gray line indicates a fourfold response ratio (near/far). Color denotes monkey (black for m39, blue for m31).

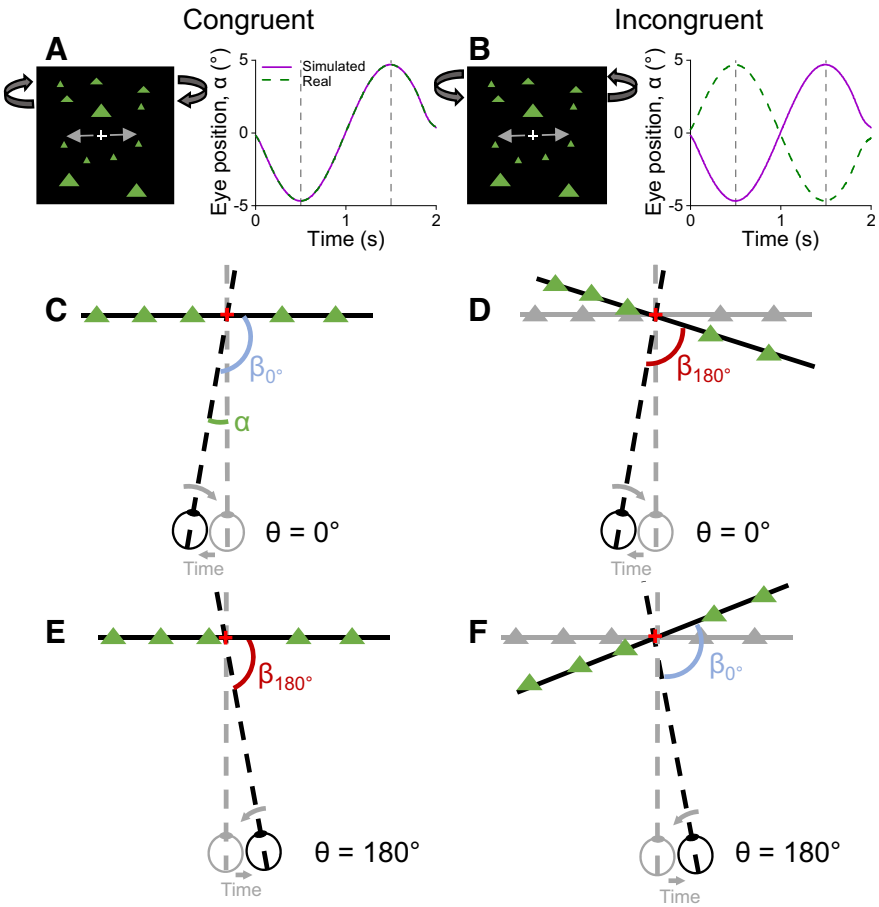


Figure 7. Schematic illustration of stimulus configurations for the Congruent (left column) and Incongruent (right column) conditions. ***A***, Illustration of the Congruent stimulus condition for a movement axis of 0°. Left, The entire scene (including fixation point) translates back and forth sinusoidally (Movie 3), while the monkey maintains fixation on the target (white cross). Right, This configuration produces dynamic perspective cues (simulated rotation, dashed green curve) that are consistent with the actual rotation of the eye (magenta curve). Dashed vertical lines, Data analysis window. ***B***, Illustration of the Incongruent condition. Left, While the scene (including fixation target) translates laterally in front of the monkey, the cloud of background triangles counterrotates around the fixation point to visually simulate eye rotation in the opposite direction. Right, This configuration produces visual cues to eye rotation (dashed green) that are in antiphase to real eye rotation (magenta). ***C***, Top-down illustration of the Congruent condition for a case in which the eye translates leftward and counterrotates rightward ($\theta = 0°$). α denotes the angular position of the eye. β_0 refers to the angle of the background relative to the line of sight. Note that, in this schematic, the background is shown as a plane of triangles for clarity of visual illustration; the actual background was a 3D cloud of triangles. ***D***, Top-down illustration of the Incongruent condition for the case of leftward eye translation ($\theta = 0°$). The background elements (green triangles) rotate around the fixation point such that the angle of the background relative to the line of sight (β) is consistent with that produced by leftward eye rotation in the Congruent condition (see ***E***). ***E***, Top-down illustration of the Congruent condition for the case of rightward eye translation and leftward eye rotation ($\theta = 180°$). Format is the same as in ***C***. ***F***, Top-down illustration of the Incongruent condition for the case of rightward eye translation and leftward eye rotation ($\theta = 180°$). Note that the angle of the background relative to the line of sight is consistent with that produced by rightward eye rotation in the Congruent condition (***C***). Format is the same as in ***D***.

roughly align with the gray dashed line, indicating a fourfold ratio of response in favor of near triangles. Furthermore, since near triangles are much larger than far triangles in the display (see Materials and Methods), this analysis may underestimate the degree to which near triangles dominate MSTd responses. Thus, it seems very unlikely that the 2D motion of far triangles could selectively drive responses of MSTd neurons in the DP condition.

Together, these findings are not consistent with the idea that responses in the DP condition are driven by basic preferences of MSTd neurons for the direction and speed of 2D motion. Rather, our data support the notion that these MSTd neurons respond selectively to the 3D rotational pattern of optic flow around the fixation point (i.e., dynamic perspective cues; see also Discussion).

Predicting responses to Congruent and Incongruent conditions with a linear model

Thus far, we have demonstrated that MSTd neurons are selective for the direction of eye rotation based on both visual and extraretinal signals. We now examine how these neurons respond to either congruent or incongruent combinations of real and simulated eye rotation, and whether we can predict these combined responses from activity measured in the EO and DP conditions.

The Congruent condition (Fig. 7*A*) is identical to the EO condition, with the addition of a stationary 3D cloud of triangles present in the background. The entire visual scene translates in front of the monkey while he pursues the fixation target (Movie 3). As the monkey pursues the target, eye orientation relative to the scene changes smoothly (Fig. 7*C*,*E*), generating rotational optic flow around the fixation point similar to that presented in the DP condition (identical if pursuit is perfect). Thus, the Congruent condition provides a combination of signals from the EO and DP conditions.

For the Incongruent condition, scene translation and the required eye movements were identical to those in the Congruent condition. The key difference is that the 3D cloud of triangles rotates around the fixation point in the opposite direction to that expected from the real eye rotation (Fig. 7*B*, Movie 4). This creates a rotational flow field that is consistent with eye movement in the opposite direction of the real eye rotation (Fig. 7, compare *D*, *E* or *C*, *F*). By creating rotational flow (dynamic perspective) that is out of phase with the real eye movement, the Incongruent condition serves to constrain the contributions of visual and extraretinal signals to the combined response.

Figure 8 shows tuning curves for an example multiunit in all four stimulus conditions. This unit has a preference for real (Fig. 8*A*) and simulated (Fig. 8*B*) eye rotation of ~90°, corresponding to an upward movement along the vertical axis (Fig. 2*A*). Responses to the Congruent condition (Fig. 8*C*) are quite similar to rotation tuning in the EO and DP conditions, as expected, given the similarity of EO and DP responses. In contrast, for the Incongruent condition, the tuning curve is clearly bimodal (Fig. 8*D*), owing to the 180° phase shift of the rotational flow field in the stimulus (Fig. 7*B*), which is captured by the phase-shifted DP tuning curve (Fig. 8*B*, gray). This pattern of results for the example unit is very well predicted by a simple weighted summation model (for details, see Materials and Methods) in which responses to the EO and DP conditions are weighted and summed to predict responses to the Congruent

and Incongruent conditions (Fig. 8*C*,*D*, red dashed curves; VAF = 0.95).

Figure 9 shows data from an example SU from MSTd, which also happens to prefer upward eye rotation (~90°) in the EO (Fig. 9*A*) and DP (Fig. 9*B*) conditions. Despite having very similar tuning curves in the EO and DP conditions, responses of this neuron showed strongly bimodal tuning in the Congruent condition (Fig. 9*E*), with a clear peak at 270° that cannot be accounted for by tuning in the EO or DP conditions. Accordingly, linear model fits (Fig. 9*E*,*F*, red dashed curves) poorly captured the response of the neuron to the Congruent and Incongruent conditions.

To understand this failure of the linear model for the unit of Figure 9, we considered that MSTd neurons are known to respond to retinal slip of a target during smooth pursuit (Komatsu and Wurtz, 1988b; Inaba et al., 2007). Thus, we computed retinal slip of the pursuit target in the Congruent and Incongruent conditions, and predicted the response of the unit to retinal slip from its direction and speed tuning for each rotation direction (for details, see Materials and Methods). The predicted responses to retinal slip show a prominent peak at 270° (Fig. 9*C*,*D*). When this retinal slip predictor was incorporated into our linear model, performance improved markedly (Fig. 9*E*,*F*, dotted blue curves), with VAF increasing from 0.48 to 0.86. Indeed, the model incorporating retinal slip predicts responses significantly better for the example neuron of Figure 9 ($p = 9.25 \times 10^{-5}$, sequential *F* test). In contrast, the example neuron of Figure 8 showed a rather flat retinal slip predictor (data not shown), such that the data were well fit by a linear model without the retinal slip predictor.

Results for the example units in Figures 8 and 9 suggest that responses in the Congruent and Incongruent conditions may reflect a simple weighted sum of responses to visual and extraretinal cues to eye rotation, with some neurons requiring a contribution from retinal slip.

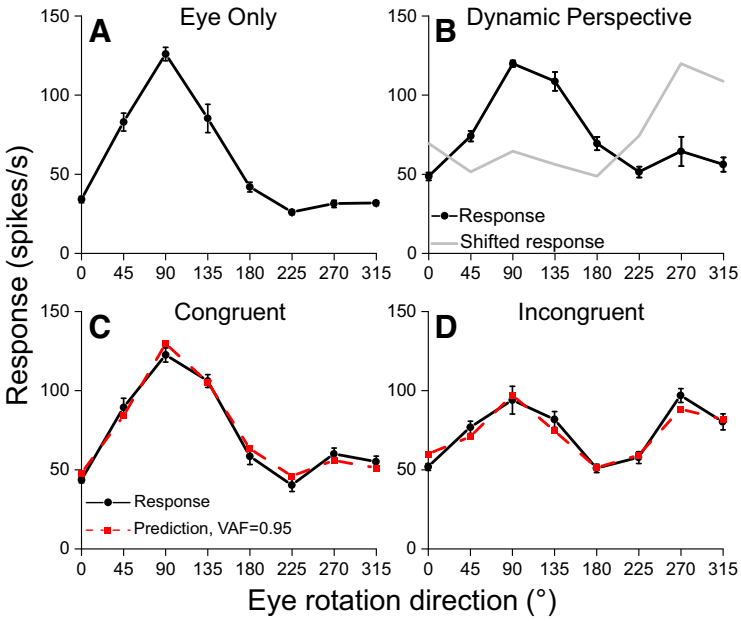


Figure 8. Responses of an example multiunit to Congruent and Incongruent conditions, along with predictions of a weighted summation model. ***A***, Rotation direction tuning of an example MU in the EO condition. ***B***, Rotation tuning of the same MU in the DP condition (black symbols). The gray curve shows the tuning curve shifted by 180° (which corresponds to a predictor of responses in the Incongruent condition). ***C***, Responses of the same example MU in the Congruent condition (black symbols, line), along with the best fit of a weighted linear summation model (red squares, dashed line). ***D***, Responses of the same MU to the Incongruent condition (black) and the best fit of the linear model (red). Note that the second response peak near 270° reflects the contribution from DP cues, which are shifted by 180° in the stimulus (***B***, gray line). Error bars in all panels represent the SEM.

Population summary of performance for the linear model

To summarize performance of our linear model across the population of MSTd neurons, we calculated the VAF by our linear model, with and without the retinal slip predictor included in the model (Fig. 10*A*). All units with significant 2D direction and speed tuning (ANOVA, $p < 0.05$) in preliminary tests (see Materials and Methods) were included. We further selected units with significant tuning for rotation in both the EO and DP conditions, as well as significant tuning in either the Congruent or Incongruent condition ($p < 0.05$, ANOVA). The linear model performed well, overall, with a median VAF value of 0.79 when the retinal slip predictor was included and a median value of 0.75 without. At the individual unit level, 18 of 97 units had a significantly better fit using the retinal slip predictor ($p < 0.05$, sequential *F* test; Fig. 10*A*, triangles). Overall, the interaction between extraretinal and visual signals regarding eye rotation is well described by a weighted linear combination. Previous literature has also shown MSTd neurons to combine other signals, such as visual and vestibular signals to heading, in a similar manner (Morgan et al., 2008).

We examined the linear model weights on EO and DP responses to gain further insight into the interactions. A negative correlation was observed between the EO and DP weights (Fig. 10*B*): if a unit has a large DP weight, it tends to have a small EO weight, and vice versa. This was true for both m39 ($R = -0.67$, $p = 6.24 \times 10^{-5}$, Spearman's rank correlation) and m31 ($R = -0.52$, $p = 3.26 \times 10^{-6}$) individually, indicating that neurons vary from DP dominant to EO dominant. The weights of most units were below unity, indicating that EO and DP signals typically combine subadditively. EO weights had a mean value of 0.394 that was significantly less than unity ($t_{(98)} = -12.8$, $p = 5.32 \times 10^{-23}$, one-sample *t* test). Similarly, DP weights had a mean value of 0.575 that was also significantly less than unity ($t_{(98)} = -17.0$, $p = 2.44 \times 10^{-31}$).

Finally, we wanted to know whether the relative weights of EO and DP responses could be explained by the relative response strengths or tuning strengths for the two isolated cues. To investigate this, we first normalized the peak–trough response differences in the EO and DP conditions by dividing them by the sum of peak–trough response differences in both conditions. Weights from the linear model fits were normalized in a similar manner, by dividing by the sum of the weights in both conditions. We found no significant correlation between normalized peak–trough responses and model weights for the EO and DP conditions for both monkeys ($p > 0.29$ for all four comparisons, Spearman's rank correlation).

We also normalized DirDI values in the same manner and compared them with model weights for the EO and DP conditions. We found a marginally significant correlation between DirDI and EO weight for m31 ($p = 0.04$, Spearman's rank correlation) but no significant correlations in the other three

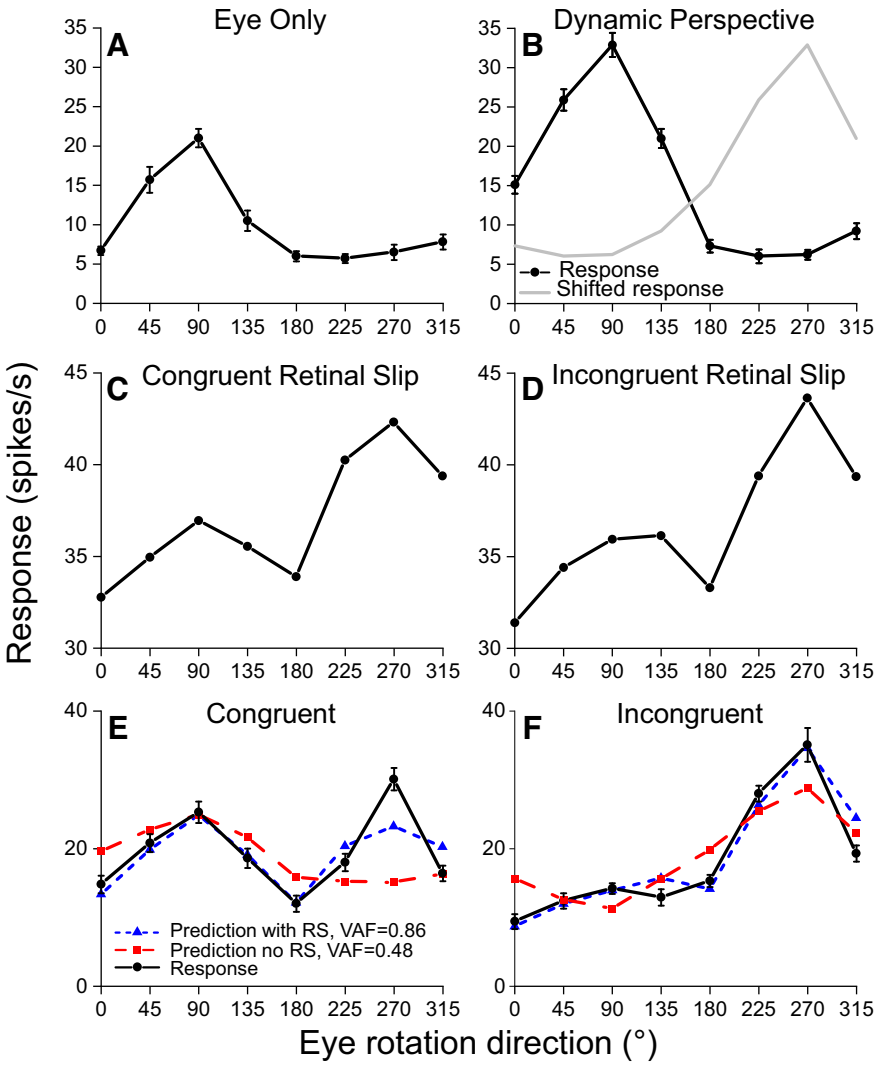


Figure 9. Predictions of weighted summation model for an example single unit with a substantial contribution from retinal slip. ***A***, ***B***, Rotation direction tuning for the EO and DP conditions, respectively. Format is the same as in Figure 8, ***A*** and ***B***. ***C***, ***D***, Predicted responses of the same example SU to retinal slip associated with eye movements in the Congruent and Incongruent conditions, respectively. See text for details regarding computation of the retinal slip predictor. ***E***, Measured (black symbols, solid line) responses of the example SU in the Congruent condition, plotted together with predictions of two versions of the weighted summation model that either includes (blue triangles, dotted lines) or does not include (red squares, dashed line) the retinal slip predictor. ***F***, Measured responses and model fits for the Incongruent condition. Format is the same as in ***E***.

comparisons ($p > 0.22$, Spearman's rank correlation). We conclude that the specific ways that individual MSTd units combine visual and extraretinal signals related to eye rotation is not well predicted from the response strengths or tuning indices measured in the EO and DP conditions.

Discussion

Our results demonstrate that many neurons in macaque area MSTd are selective for rotational components of optic flow produced by eye rotations relative to the scene, and that this visual selectivity for eye rotation is typically congruent with rotation tuning based on extraretinal signals. Responses to visually simulated eye rotation in the DP condition cannot be explained by the standard 2D direction and speed tuning of these neurons, indicating that they respond to global rotational flow cues around the point of fixation. This form of rotation tuning in MSTd is distinct from that described previously for rotations around other axes, as discussed below. Rotation preferences are also clustered in MSTd, suggesting a functional organization for representing eye rotation based on both visual and extraretinal signals. Finally, responses to combinations of real and visually simulated eye rotations are well predicted by a simple linear summation model, analogous to the integration of visual and vestibular signals in MSTd (Morgan et al., 2008). Our findings provide the first systematic evidence of a neural substrate for representing eye rotations based on rotational flow cues, and suggest that MSTd constructs a unified representation of eye rotations that integrates both visual and extraretinal signals.

Roles of MSTd in representing smooth pursuit eye movements

MSTd has long been suggested to play important roles in smooth pursuit eye movements, being important for volitional pursuit (Ono and Mustari, 2006), for maintaining pursuit (Newsome et al., 1988), and for modulating pursuit gain (Churchland and Lisberger, 2002, 2005). Our findings from the EO condition are broadly consistent with previous work, indicating that a large fraction of MSTd neurons carry extraretinal signals related to pursuit eye movements.

The key novelty in our study involves demonstrating that area MSTd also carries visual signals regarding smooth eye movements, driven by rotational optic flow (i.e., dynamic perspective) cues. Recent work has provided a growing body of evidence that visual signals regarding eye rotation play key roles in computations that require knowledge of smooth eye velocity; this includes computing heading during eye movements (Sunkara et al., 2015; Manning and Britten, 2019; Danz et al., 2020) and computing depth from motion parallax (Kim et al., 2015, 2017; Xu and DeAngelis, 2022). While these studies have emphasized the involvement of visual cues to eye rotation, the origin of these visual rotation signals has remained unclear. Our data reveal MSTd as a structure that is well suited for processing dynamic perspective cues to eye rotation and integrating them with extraretinal signals regarding eye movement.

Why might the brain want to use visual cues to eye rotation when extraretinal signals are also available? One key reason is that dynamic perspective cues directly provide information about eye rotation relative to the scene. In contrast, calculating eye rotation relative to the scene from extraretinal signals is more complex when eye, head, and body may all rotate simultaneously. In that case, extraretinal signals regarding eye-in-head, head-on-body, and body-in-world movements may all need to be integrated to compute eye movement relative to the scene. For

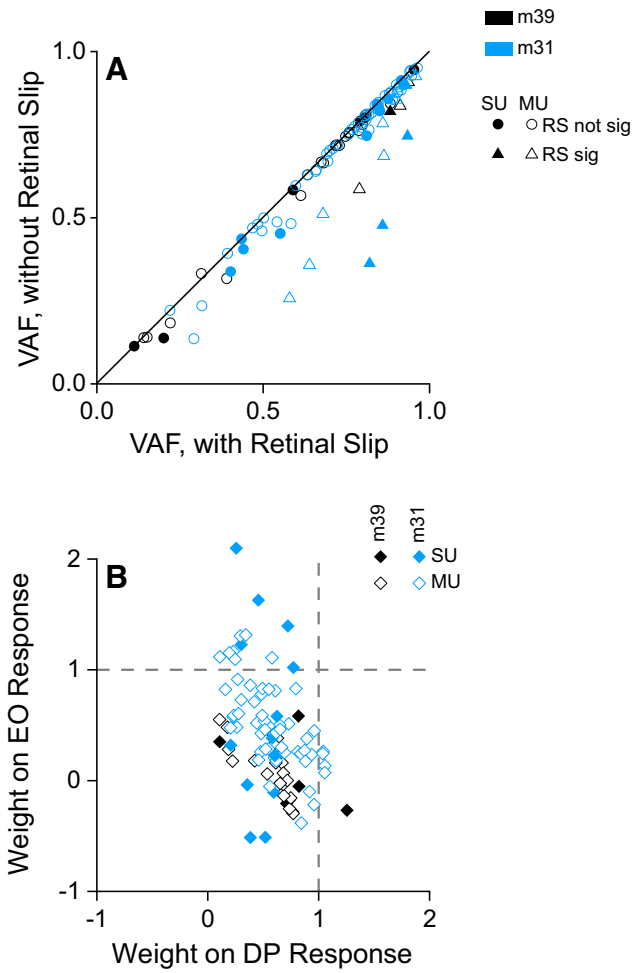


Figure 10. Population summary of linear model performance. ***A***, VAF without the RS factor is plotted against VAF with the RS factor for m39 (black) and m31 (blue). Open and filled symbols denote MU and SU activity, respectively. Triangles denote units for which the retinal slip factor significantly improves model predictions (7 of 29 units for m39; 11 of 68 for m31; sequential *F* test, $p < 0.05$), whereas circles denote units for which the retinal slip contribution is not significant. Units with significant tuning for basic protocols (2D motion and speed), significant tuning for both EO and DP conditions, and significant tuning for (at least) one of the combined conditions were included. ***B***, Model weights on responses to the EO condition are plotted against weights on responses to the DP condition for m39 (black) and m31 (blue). Data are shown for the same populations of units as in ***A***. Open and filled symbols denote MU and SU activity, respectively.

some computations that require information about eye movements relative to the scene, such as computation of depth from motion parallax (Nawrot, 2003; Nadler et al., 2009; Nawrot and Stroyan, 2009; Kim et al., 2015, 2016), it could be more efficient to rely on visual rotation cues than extraretinal signals. In general, the brain is most likely to make use of both visual and extraretinal signals regarding eye rotation when both are available.

Related to this point, it is worth cautioning that the contribution of visual rotation cues can be easily mistaken for an effect of extraretinal signals in studies where there is a visible stationary background, unless considerable care is taken to dissociate the two signals (Manning and Britten, 2019). Thus, it is possible that some effects previously attributed to extraretinal signals may have instead been driven by visual cues to eye rotation.

Relationship to other forms of visual selectivity in MSTd

Neurons in area MSTd are known to respond to various forms of image motion; thus, it is critical to address how our findings are distinct from other forms of 2D motion and rotation selectivity described in the literature. Consistent with most previous studies of pursuit responses in MSTd (Komatsu and Wurtz, 1988b; Ono and Mustari, 2006; Ono et al., 2010), we found that direction preferences for 2D motion and eye rotation were typically opposite. One exception to this tendency was the study of Squatrito and Maioli (1997), which found completely aligned preferences between 2D motion and real eye rotation. The reasons for this discrepancy are not clear, but it is worth noting that only 18 of 132 neurons in that study had significant direction tuning for both 2D visual motion and smooth pursuit (Squatrito and Maioli, 1997). This exception notwithstanding, the general finding of opposite preferences for 2D visual motion and real eye rotation is not compatible with the possibility that responses in our DP condition are driven by 2D visual motion selectivity. As discussed in the text related to Figure 6, responses in the DP condition could only be explained by 2D motion selectivity if responses were dominated by far triangles that move slowly and are much smaller in the image. Our analysis suggests that this is implausible (Fig. 6*D*,*E*). Thus, we conclude that tuning in our DP condition reflects a form of selectivity for visual rotation around the fixation point.

How is tuning in our DP condition related to other forms of rotational selectivity described in area MSTd? One type of rotational optic flow that has been studied in MSTd involves circular motion in the image plane (Tanaka and Saito, 1989; Duffy and Wurtz, 1991b), sometimes combined with fore/aft motion in what is known as a spiral space (Graziano et al., 1994). While this type of circular optic flow may be produced by a roll rotation of the head/eye, it is very different from the rotational pattern produced by the combinations of eye translation and rotation that we study (Movie 2). Another form of rotational optic flow that has been studied in MSTd involves pure rotations of the eye/head around the center of the head, which has also been studied in combination with physical rotation of the head/body (Takahashi et al., 2007). This type of rotation produces a flow field in which all background elements move in the same direction and retinal velocity is independent of depth, which is again quite different from the flow fields studied here.

Thus, to our knowledge, no previous study has systematically examined MSTd responses to flow fields that simulate the common ecological situation in which a translating observer counter-rotates their eyes to maintain fixation on a world-fixed target during self-motion (Fig. 1). Interestingly, one of the earliest studies of MSTd neurons (Saito et al., 1986) may have identified a small population of neurons with DP selectivity. Saito et al. (1986, their Fig. 10) presented monkeys with a textured flat board that rotated (in depth) around an axis in the frontoparallel plane. They identified 17 neurons that were selective for the direction of rotation in depth ("Rd" cells) and had varying preferred axes of rotation (although data from only one example neuron are shown). The authors attributed these findings to the combination of expansion and contraction in receptive fields and did not consider that this selectivity might be related to representing eye movements. It seems likely to us that responses of these Rd cells were being modulated by dynamic perspective cues.

Finally, we note that the blank background in our EO condition is not truly dark, as the backlighting of the projector leaves a faintly visible fine texture on the display. Could responses of MSTd neurons in our EO condition have been driven by this faint background texture, as observed under some conditions in a previous study of MSTd neurons (Chowdhury et al., 2009)?

Since the background texture is static on the display, eye movements in the EO condition would produce full-field motion of the faint texture in the direction opposite to the eye movement. It is quite unlikely that this background motion accounts for EO responses, since the pattern of motion of this faint texture would be very different from the pattern of image motion presented in the DP condition. Specifically, image motion of any faint background texture in the EO condition would lack the critical motion parallax cues that simulate eye rotation in the DP condition. Given that rotation tuning preferences in the EO and DP conditions are closely matched for most neurons, rotation tuning in the EO condition would thus be very difficult to explain by image motion of the faint background texture.

Future directions

We observed that selectivity for rotation direction in the EO and DP conditions is generally matched for simultaneously recorded single neurons and multiunits. This strongly suggests a clustered representation of eye rotation in area MSTd, consistent with what has been observed previously for heading tuning based on optic flow (Britten, 1998; Chen et al., 2008). This clustering may be indicative of a topographic map of eye velocity, although we cannot infer this with our methods. Fortunately, clustering of rotation selectivity enables microstimulation experiments that should be able to test whether MSTd makes a causal contribution to computations that rely on estimates of eye velocity during self-motion. For example, it would be interesting to test whether microstimulation of area MSTd could induce depth-sign selectivity in area MT when the visual stimulus is otherwise depth sign ambiguous (Nadler et al., 2008, 2009). Along the same lines, it would be interesting to reversibly inactivate area MSTd and examine whether depth-sign selectivity in area MT is attenuated.

Going forward, it is critical that studies of visual processing during pursuit eye movements account properly for dynamic perspective cues associated with eye rotation. Some previous studies have simulated pursuit eye movements using laminar image motion (Bradley et al., 1996; Shenoy et al., 1999). This is not an accurate simulation of eye rotation because it lacks dynamic perspective cues. Hence, these studies likely misestimated the relative contributions of extraretinal and visual cues involved in pursuit compensation (Sunkara et al., 2015). Our findings establish area MSTd as a neural substrate for representing smooth eye movements during self-motion based on both visual and extraretinal signals, thus opening the door to exploring the circuit mechanisms by which information about eye rotation combines with other signals to perform a variety of useful computations.

Citation diversity statement

Recent work in neuroscience and other fields has drawn attention to citation biases such that women and other minorities are undercited (Dworkin et al., 2020; Zurn et al., 2020). To bring awareness to the issue, we have chosen to include a citation diversity statement and report the statistics of author gender in our citations. Although we recognize that this method has limitations (Zurn et al., 2020), we are committed to acknowledging these biases and being transparent in the demographics of the authors we cite. In neuroscience, top journals reported 58.6% man/man, 25.3% woman/man, 9.4% man/woman, and 6.7% woman/woman as first/last authors. Our references contain 67.5% man/man, 18.2% woman/man, 9.1% man/woman, and 5.2% woman/woman as first/last authors.

References

Albright TD, Desimone R (1987) Local precision of visuotopic organization in the middle temporal area (Mt) of the macaque. Exp Brain Res 65:582–592.

Albright TD, Desimone R, Gross CG (1984) Columnar organization of directionally selective cells in visual area MT of the macaque. J Neurophysiol 51:16–31.

Banks MS, Ehrlich SM, Backus BT, Crowell JA (1996) Estimating heading during real and simulated eye movements. Vision Res 36:431–443.

Ben Hamed S, Page W, Duffy C, Pouget A (2003) MSTd neuronal basis functions for the population encoding of heading direction. J Neurophysiol 90:549–558.

Bradley DC, Maxwell M, Andersen RA, Banks MS, Shenoy KV (1996) Mechanisms of heading perception in primate visual cortex. Science 273:1544–1547.

Britten KH (1998) Clustering of response selectivity in the medial superior temporal area of extrastriate cortex in the macaque monkey. Vis Neurosci 15:553–558.

Burlingham CS, Heeger DJ (2020) Heading perception depends on time-varying evolution of optic flow. Proc Natl Acad Sci U S A 117:33161–33169.

Chen A, Gu Y, Takahashi K, Angelaki DE, DeAngelis GC (2008) Clustering of self-motion selectivity and visual response properties in macaque area MSTd. J Neurophysiol 100:2669–2683.

Chowdhury SA, Takahashi K, DeAngelis GC, Angelaki DE (2009) Does the middle temporal area carry vestibular signals related to self-motion? J Neurosci 29:12020–12030.

Chukoskie L, Movshon JA (2009) Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. J Neurophysiol 102:3225–3233.

Churchland AK, Lisberger SG (2002) Gain control in human smooth-pursuit eye movements. J Neurophysiol 87:2936–2945.

Churchland AK, Lisberger SG (2005) Relationship between extraretinal component of firing rate and eye speed in area MST of macaque monkeys. J Neurophysiol 94:2416–2426.

Crowell JA, Banks MS, Shenoy KV, Andersen RA (1998) Visual self-motion perception during head turns. Nat Neurosci 1:732–737.

Danz AD, Angelaki DE, DeAngelis GC (2020) The effects of depth cues and vestibular translation signals on the rotation tolerance of heading tuning in macaque area MSTd. eNeuro 7:ENEURO.0259-20.2020.

DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. J Neurophysiol 89:1094–1111.

Desimone R, Ungerleider LG (1986) Multiple visual areas in the caudal superior temporal sulcus of the macaque. J Comp Neurol 248:164–189.

Duffy CJ, Wurtz RH (1991a) Sensitivity of Mst neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. J Neurophysiol 65:1346–1359.

Duffy CJ, Wurtz RH (1991b) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. J Neurophysiol 65:1329–1345.

Duffy CJ, Wurtz RH (1995) Response of monkey Mst neurons to optic flow stimuli with shifted centers of motion. J Neurosci 15:5192–5208.

Dworkin JD, Linn KA, Teich EG, Zurn P, Shinohara RT, Bassett DS (2020) The extent and drivers of gender imbalance in neuroscience reference lists. Nat Neurosci 23:918–926.

Fetsch CR, Wang S, Gu Y, DeAngelis GC, Angelaki DE (2007) Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. J Neurosci 27:700–712.

Gattass R, Gross CG (1981) Visual topography of striate projection zone (Mt) in posterior superior temporal sulcus of the macaque. J Neurophysiol 46:621–638.

Graziano MSA, Andersen RA, Snowden RJ (1994) Tuning of Mst neurons to spiral motions. J Neurosci 14:54–67.

Grigo A, Lappe M (1999) Dynamical use of different sources of information in heading judgments from retinal flow. J Opt Soc Am A 16:2079–2091.

Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 26:73–85.

Ilg UJ, Thier P (2003) Visual tracking neurons in primate area MST are activated by smooth-pursuit eye movements of an "imaginary" target. J Neurophysiol 90:1489–1502.

Inaba N, Shinomoto S, Yamane S, Takemura A, Kawano K (2007) MST neurons code for visual motion in space independent of pursuit eye movements. J Neurophysiol 97:3473–3483.
Inaba N, Miura K, Kawano K (2011) Direction and speed tuning to visual motion in cortical areas MT and MSTd during smooth pursuit eye movements. J Neurophysiol 105:1531–1545.
Kim HR, Angelaki DE, DeAngelis GC (2015) A novel role for visual perspective cues in the neural computation of depth. Nat Neurosci 18:129–137.
Kim HR, Angelaki DE, DeAngelis GC (2016) The neural basis of depth perception from motion parallax. Philos Trans R Soc Lond B Biol Sci 371:20150256.
Kim HR, Angelaki DE, DeAngelis GC (2017) Gain modulation as a mechanism for coding depth from motion parallax in macaque area MT. J Neurosci 37:8180–8197.
Koenderink JJ, van Doorn AJ (1976) Local structure of movement parallax of the plane. J Opt Soc Am 66:717–723.
Komatsu H, Wurtz RH (1988a) Relation of cortical areas Mt and Mst to pursuit eye-movements.1. localization and visual properties of neurons. J Neurophysiol 60:580–603.
Komatsu H, Wurtz RH (1988b) Relation of cortical areas Mt and Mst to pursuit eye-movements. 3. interaction with full-field visual-stimulation. J Neurophysiol 60:621–644.
Krauzlis RJ, Lisberger SG (1996) Directional organization of eye movement and visual signals in the floccular lobe of the monkey cerebellum. Exp Brain Res 109:289–302.
Krauzlis RJ, Miles F (1996) Decreases in the latency of smooth pursuit and saccadic eye movements produced by the "gap paradigm" in the monkey. Vision Res 36:1973–1985.
Li L, Warren WH (2000) Perception of heading during rotation: sufficiency of dense motion parallax and reference objects. Vision Res 40:3873–3894.
Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. Proc R Soc Lond B Biol Sci 208:385–397.
Manning TS, Britten KH (2019) Retinal stabilization reveals limited influence of extraretinal signals on heading tuning in the medial superior temporal area. J Neurosci 39:8064–8078.
Maunsell JHR, Van Essen DC (1983) The connections of the middle temporal visual area (Mt) and their relationship to a cortical hierarchy in the macaque monkey. J Neurosci 3:2563–2586.
Morgan ML, DeAngelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. Neuron 59:662–673.
Nadler JW, Angelaki DE, DeAngelis GC (2008) A neural representation of depth from motion parallax in macaque visual cortex. Nature 452:642–645.
Nadler JW, Nawrot M, Angelaki DE, DeAngelis GC (2009) MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. Neuron 63:523–532.
Naji JJ, Freeman TCA (2004) Perceiving depth order during pursuit eye movement. Vision Res 44:3025–3034.
Nawrot M (2003) Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. Vision Res 43:1553–1562.
Nawrot M, Stroyan K (2009) The motion/pursuit law for visual depth perception from motion parallax. Vision Res 49:1969–1978.
Newsome WT, Wurtz RH, Komatsu H (1988) Relation of cortical areas Mt and Mst to pursuit eye-movements. II. Differentiation of retinal from extraretinal inputs. J Neurophysiol 60:604–620.
Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J Neurosci 25:10049–10060.
Ono S, Mustari MJ (2006) Extraretinal signals in MSTd neurons related to volitional smooth pursuit. J Neurophysiol 96:2819–2825.
Ono S, Brostek L, Nuding U, Glasauer S, Büttner U, Mustari MJ (2010) The response of MSTd neurons to perturbations in target motion during ongoing smooth-pursuit eye movements. J Neurophysiol 103:519–530.
Prince SJ, Pointon AD, Cumming BG, Parker AJ (2002) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J Neurophysiol 87:191–208.
Rieger JH, Lawton DT (1985) Processing differential image motion. J Opt Soc Am A 2:354–360.
Robinson DA (1965) The mechanics of human smooth pursuit eye movement. J Physiol 180:569–591.
Rodman HR, Albright TD (1987) Coding of visual stimulus velocity in area MT of the macaque. Vision Res 27:2035–2048.
Roy JP, Komatsu H, Wurtz RH (1992) Disparity sensitivity of neurons in monkey extrastriate area MST. J Neurosci 12:2478–2492.
Royden CS (1997) Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. J Opt Soc Am A 14:2128–2143.
Royden CS, Banks MS, Crowell JA (1992) The perception of heading during eye-movements. Nature 360:583–585.
Royden CS, Crowell JA, Banks MS (1994) Estimating heading during eye-movements. Vision Res 34:3197–3214.
Saito H, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. J Neurosci 6:145–157.
Shao MM, DeAngelis GC, Angelaki DE, Chen AH (2018) Clustering of heading selectivity and perception-related activity in the ventral intraparietal area. J Neurophysiol 119:1113–1126.
Shenoy KV, Bradley DC, Andersen RA (1999) Influence of gaze rotation on the visual response of primate MSTd neurons. J Neurophysiol 81:2764–2786.
Squatrito S, Maioli MG (1997) Encoding of smooth pursuit direction and eye position by neurons of area MSTd of macaque monkey. J Neurosci 17:3847–3860.
Sunkara A, DeAngelis GC, Angelaki DE (2015) Role of visual and non-visual cues in constructing a rotation-invariant representation of heading in parietal cortex. Elife 4:e04693.
Sunkara A, DeAngelis GC, Angelaki DE (2016) Joint representation of translational and rotational components of optic flow in parietal cortex. Proc Natl Acad Sci U S A 113:5077–5082.
Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE (2007) Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. J Neurosci 27:9742–9756.
Takemura A, Inoue Y, Kawano K (2000) The effect of disparity on the very earliest ocular following responses and the initial neuronal activity in monkey cortical area MST. Neurosci Res 38:93–101.
Tanaka K, Saito HA (1989) Analysis of motion of the visual-field by direction, expansion contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. J Neurophysiol 62:626–641.
Tanaka K, Sugita Y, Moriya M, Saito HA (1993) Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual-cortex. J Neurophysiol 69:128–142.
Ungerleider LG, Desimone R (1986) Cortical connections of visual area Mt in the macaque. J Comp Neurol 248:190–222.
Van Essen DC, Maunsell JHR, Bixby JL (1981) The middle temporal visual area in the macaque: myeloarchitecture, connections, functional-properties and topographic organization. J Comp Neurol 199:293–326.
Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc 8:443–459.
von Holst E, Mittelstaedt H (1950) Das reafferenzprinzip: wechselwirkungen zwischen zentralnervensystem und peripherie. Naturwissenschaften 37:464–476.
Wallach H (1987) Perceiving a stable environment when one moves. Annu Rev Psychol 38:1–27.
Xu ZX, DeAngelis GC (2022) Neural mechanism for coding depth from motion parallax in area MT: gain modulation or tuning shifts? J Neurosci 42:1235–1253.
Yang Y, Liu S, Chowdhury SA, DeAngelis GC, Angelaki DE (2011) Binocular disparity tuning and visual-vestibular congruency of multisensory neurons in macaque parietal cortex. J Neurosci 31:17905–17916.
Zeki SM (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573.
Zurn P, Bassett DS, Rust NC (2020) The citation diversity statement: a practice of transparency, a way of life. Trends Cogn Sci 24:669–672.