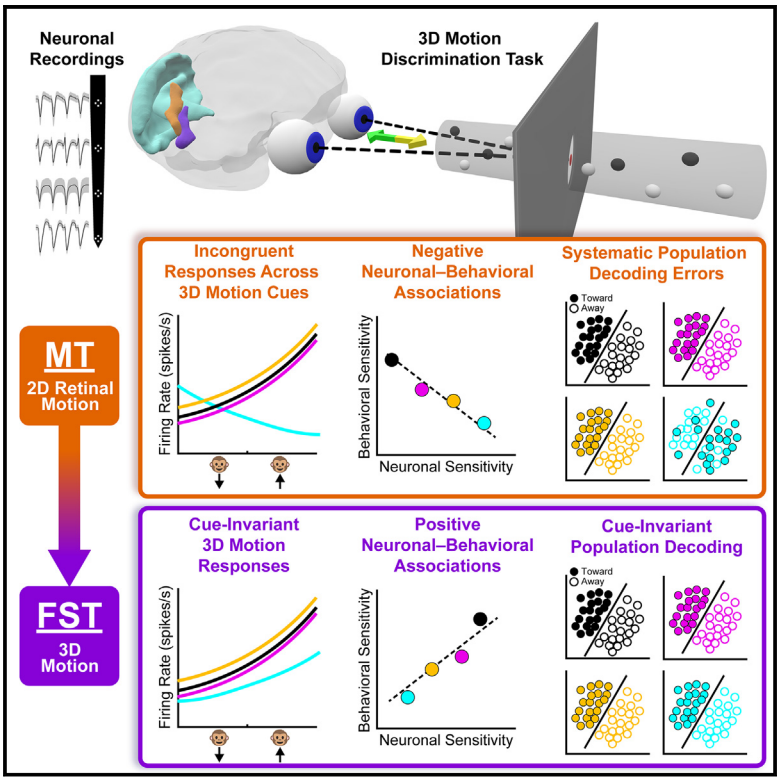
Cell Reports

Article

# Hierarchical computation of 3D motion across macaque areas MT and FST

## Graphical abstract




## Highlights

- Retinal motion is transformed into 3D motion across an MT-to-FST hierarchy
- Selective integration of MT responses explains the 3D selectivity of FST neurons
- This hierarchical transformation generalizes the decoding of 3D motion
- Neuronal activity in FST accounts for behavioral sensitivity to 3D motion


## Authors

Lowell W. Thompson, Byounghoon Kim, Bas Rokers, Ari Rosenberg

## Correspondence

ari.rosenberg@wisc.edu



## In brief

Thompson et al. investigate how 2D retinal-motion signals are hierarchically transformed into representations of 3D motion across the macaque visual cortex. The results show that integrating the output of 2D retinal-motion selective MT neurons can produce behaviorally relevant 3D motion representations in FST.


Thompson et al., 2023, Cell Reports *42*, 113524
December 26, 2023 © 2023 The Author(s).
https://doi.org/10.1016/j.celrep.2023.113524

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

# Hierarchical computation of 3D motion across macaque areas MT and FST

Lowell W. Thompson,[1,3] Byounghoon Kim,[1] Bas Rokers,[2] and Ari Rosenberg[1,4,*]
[1]Department of Neuroscience, School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI 53705, USA
[2]Department of Psychology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[3]Present address: Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Lead contact
*Correspondence: ari.rosenberg@wisc.edu
https://doi.org/10.1016/j.celrep.2023.113524

## SUMMARY

Computing behaviorally relevant representations of three-dimensional (3D) motion from two-dimensional (2D) retinal signals is critical for survival. To ascertain where and how the primate visual system performs this computation, we recorded from the macaque middle temporal (MT) area and its downstream target, the fundus of the superior temporal sulcus (area FST). Area MT is a key site of 2D motion processing, but its role in 3D motion processing is controversial. The functions of FST remain highly underexplored. To distinguish representations of 3D motion from those of 2D retinal motion, we contrast responses to multiple motion cues during a motion discrimination task. The results reveal a hierarchical transformation whereby many FST but not MT neurons are selective for 3D motion. Modeling results further show how generalized, cue-invariant representations of 3D motion in FST may be created by selectively integrating the output of 2D motion selective MT neurons.

## INTRODUCTION

The neural computation of three-dimensional (3D) motion is essential for functions such as capturing prey or avoiding predators. Although 3D visual processing shaped the evolution of the primate brain,[1] current understanding of 3D motion processing remains surprisingly limited compared with that of two-dimensional (2D) retinal-motion processing.[2–4] In part, this reflects that distinct cues signal 2D versus 3D motion, with 3D motion computations depending critically on the relationship between retinal-motion signals sensed by the two eyes.[5–8]

The inherent covariation of an object's motion in the world and the corresponding retinal projections makes dissociating 3D from 2D representations challenging. The macaque middle temporal (MT) area fundamentally contributes to the processing of 2D retinal motion.[9] Some early studies implicated MT in the processing of 3D motion,[10,11] but other work suggested that the apparent 3D selectivity was explained by lower-level selectivity for 2D motion, speed, and horizontal disparity.[12] Subsequent studies supported the possibility of selectivity for stereoscopically defined 3D motion in MT,[13,14] whereas other work isolated that selectivity to the neighboring fundus of the superior temporal sulcus (area FST).[15] Thus, the site of 3D-motion processing and the specific role of MT remains controversial.

Computational work has further suggested that the interaction of 2D direction selectivity and ocular dominance (OD) in MT may form the basis of a "non-canonical" mechanism for encoding and decoding 3D motion.[16,17] However, the ability to decode information from a population of neurons does not imply that those neurons explicitly represent that information. For instance, the information required to decode facial expression is present in the primary visual cortex (V1),[18] but V1 neurons do not represent facial expression. Indeed, 2D direction selectivity and OD are common properties of early visual system neurons. If these criteria were sufficient to define 3D motion selectivity, then 2D direction selective retinal ganglion cells would also need to be classified as 3D selective. These observations suggest that more stringent criteria are required to define 3D motion selectivity.

To clarify the contributions of MT and FST to motion processing, we contrasted neuronal responses to a set of stimuli designed to unambiguously distinguish 3D from 2D selectivity.[8] For the vast majority of MT neurons, apparent 3D direction preferences switched depending on which eye viewed the stimuli. When both eyes saw the stimuli, the responses typically followed those of the dominant eye. These findings were readily explained by a combination of 2D direction selectivity and OD. In contrast, the 3D direction preferences of FST neurons were often independent of which eye saw the stimulus and the class of motion cue. Moreover, the sensitivity of 3D motion selective FST neurons predicted the behavioral sensitivity of the monkeys in a toward or away object motion discrimination task. Modeling results further showed how generalized, cue-invariant 3D motion selectivity could be achieved by selectively integrating the output of 2D retinal-motion selective neurons. The findings collectively reveal that MT is largely specialized for processing 2D retinal motion whereas FST contains a prominent representation of 3D motion, and show how behaviorally relevant motion representations can be computed from ambiguous retinal signals.


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

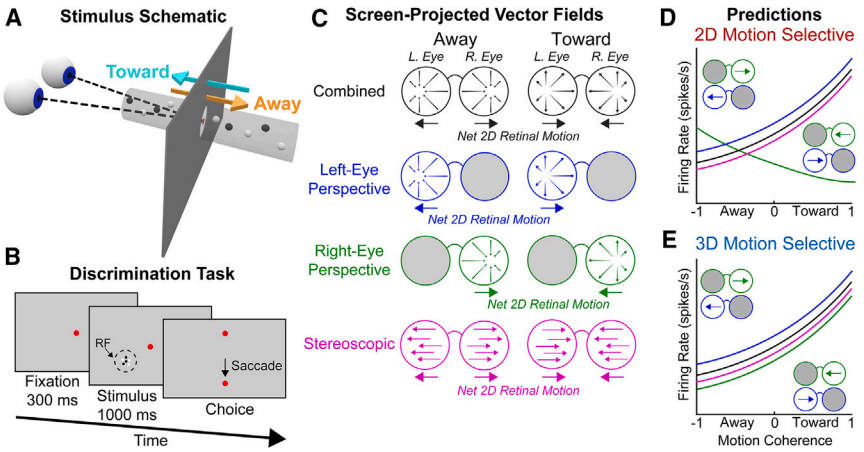



**Figure 1. Three-dimensional-motion stimuli, discrimination task, and predictions**

(A) The stimulus was a 3D volume with dots that moved toward (cyan arrow) or away (orange arrow) from the monkey's cyclopean eye.

(B) On each trial, the monkey fixated for 300 ms before a 3D motion stimulus appeared for 1 s. The perceived direction of motion was indicated by a downward ("toward") or upward ("away") saccade. Stimuli were always presented within the neuron's receptive field (RF).

(C) Screen-projected vector fields depicting the left- and right-eye signals for each cue condition (combined, black; left-eye perspective, blue; right-eye perspective, green; stereoscopic, magenta). Note that the 3D motion stimuli produced opposite net 2D retinal motions in the two eyes.

(D) Predicted responses for 2D retinal-motion selective neurons. Left- and right-eye perspective cue responses are negatively correlated. Combined-cue and stereoscopic cue responses follow the dominant eye responses.

(E) Predicted responses for 3D motion selective neurons. Preferences are the same across cue conditions.

## RESULTS

### Differentiating 3D from 2D motion selectivity

Distinguishing genuine 3D motion selectivity from 2D retinal-motion selectivity is challenging because the two signals necessarily covary. To this end, we developed a set of 3D motion stimuli consisting of dots that moved toward or away from the monkey's "cyclopean eye" (the midpoint between the two eyes) within a spatially confined volume[6–8] (STAR Methods; Figure 1A). In contrast to previous studies that assessed 3D motion selectivity in anesthetized or passively fixating monkeys,[10,12–14] we measured neuronal responses to these stimuli as the monkeys performed a toward or away motion discrimination task in which motion coherence was varied (Figure 1B). On each trial, the direction of motion was reported as either "toward" or "away." The stimulus set included four conditions across which different visual cues signaled the direction of 3D motion: distinct left-eye and right-eye perspective cues (optic flow/looming), stereoscopic cues, and all cues combined (combined-cues) (Figure 1C).

For motion toward or away from the cyclopean eye, projective geometry maximizes the velocity differences between the left- and the right-eye retinal signals, such that the net 2D retinal motions are equal and opposite in the two eyes. Leveraging this geometric property, we compared neuronal responses across the four cue conditions to distinguish 2D retinal-motion selectivity (Figure 1D) from 3D motion selectivity (Figure 1E), as the following example illustrates. Consider a neuron that responds more strongly to toward than to away motion in the combined-cue and stereoscopic cue conditions (i.e., binocular viewing conditions), as measured in previous studies.[13,14] These responses could reflect either: (1) genuine 3D motion selectivity or (2) a combination of 2D retinal-motion selectivity and OD. In particular, a 2D selective neuron with OD will preferentially respond to 3D motion that results in the dominant eye seeing the preferred 2D motion. In this example, a rightward motion selective neuron with left-eye OD will respond best to toward motion during binocular viewing (Figure 1D). Thus, responses to combined-cue and/or stereoscopic cue stimuli alone cannot distinguish 3D from 2D motion selectivity.

These alternatives can instead be distinguished by also considering responses to left- and right-eye perspective cues (i.e., monocular viewing conditions). Most notably, for 2D selective neurons, apparent 3D motion preferences will be opposite for these conditions (Figure 1D). In addition, for 2D selective neurons with strong OD, responses to combined-cue and stereoscopic cue stimuli will positively (negatively) correlate with the dominant (non-dominant) eye's perspective cue responses. For those with weak OD, responses to the binocular conditions will show little-to-no preference for either toward or away motion. In stark contrast, the preferences of 3D selective neurons will be cue invariant (Figure 1E). These predictions provide rigorous criteria for classifying neurons as selective for 2D retinal motion or 3D motion.

### Localization of MT and FST

To investigate the hierarchical transformation of visual motion, we recorded from 358 MT (monkey J, n = 193; monkey C, n = 165) and 534 FST (monkey J, n = 317; monkey C, n = 217) neurons in two male rhesus macaques using laminar array probes with either four or eight tetrodes (Figure 2A). While the functional and anatomical organization of MT is well documented,[9] little is known about FST. We therefore first identified MT, where receptive fields (RFs) are organized along a peripheral-to-foveal visual field representation moving anterior to posterior[19] (Figure 2B, left). We then localized FST, which is anteroventral to MT. At the boundary, RFs became larger and showed a coarse reversal in retinotopy. Moving farther anterior, they increased in size with substantial variability[20–22] (Figure 2B, right). We also observed that RF size increased more slowly as a function of eccentricity in MT (type II regression slope = 0.82) than FST (slope = 1.15), as indicated by a linear mixed-effects model, which showed a significant interaction between area and eccentricity (monkey as a random effect, $p = 4.80 \times 10^{-4}$). Importantly, the eccentricities at which the stimuli were presented did not significantly differ for the two areas (Wilcoxon rank sum, p = 0.17). The areas

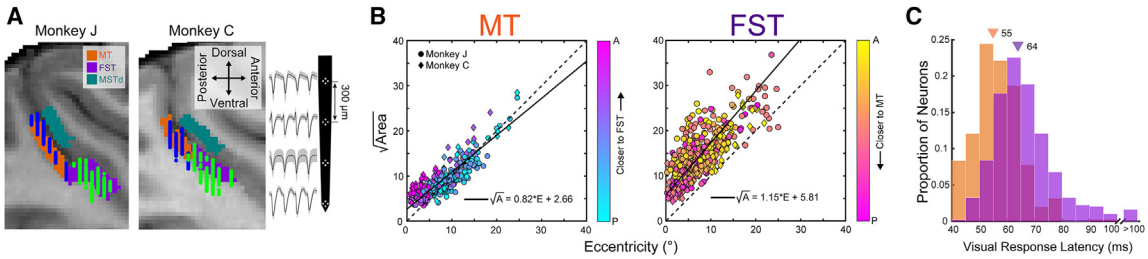



**Figure 2. Localization of MT and FST**

(A) Sagittal MRI sections from monkey J (left) and monkey C (right) showing MRI-based estimates of MT (orange), FST (purple), and the dorsal medial superior temporal (MSTd) area (teal) for reference. Blue and green dots show approximate recording locations in MT and FST, respectively. Recording sites were projected along the medial-lateral axis onto individual sections. A schematic of a laminar probe with four tetrodes and example spike waveforms is shown to the right.

(B) MT neurons (left) had smaller RFs, whose area increased at a slower rate with eccentricity compared with FST neurons (right). Type II regression lines (solid black) and identity lines (dashed black) are shown for reference. Points are colored according to the anterior-posterior (A-P) location of the recording site (normalized for each monkey). The representation of the visual field in MT progressed from posterior (larger peripheral RFs; cyan) to anterior (smaller foveal RFs; magenta). A coarse reversal occurred in FST, where posterior RFs were more foveal (magenta) and anterior RFs were more peripheral (yellow).

(C) Distribution of visual response latencies in MT (orange) and FST (purple). See also Figure S1.

were further distinguished by their visual response latencies, which were significantly shorter in MT (median = 55 ms) than in FST (median = 64 ms; Wilcoxon rank sum, $p = 4.19 \times 10^{-40}$; Figure 2C). There was not a difference, however, in whether the neurons preferred 2D motion presented at a slower (4.2°/s) or faster (12.6°/s) speed. For neurons tested at both speeds, the proportions that preferred the slow speed were similar in the two areas (MT, 24%, n = 40/165; FST, 31%, n = 162/523) and not significantly different (chi-squared test, p = 0.10). Last, differences in saccade-related activity during an overlap saccade task also distinguished the areas (Figure S1A; STAR Methods). In MT, only a small proportion of neurons showed increases in activity leading up to the saccade (n = 35/358, 10%; Figure S1B). The prevalence of saccade-related activity in FST was nearly double that in MT (n = 98/534, 18%) and the modulation was more pronounced (Figure S1C). Thus, multiple physiological properties distinguished the areas and provide systematic criteria for localizing FST.

## MT and FST respond differently to 3D motion

To identify neurons that were putatively selective for 3D motion, we performed two-way ANOVAs over motion direction and coherence on the responses to the combined-cue stimuli, which include all of the 3D-motion cues and elicit the most precise estimates of 3D motion[8] (Figure 1C). The proportions of neurons with statistically significant tuning ($p < 0.05$ for the main effect of direction) were similar in MT (54%, n = 194/358) and FST (58%, n = 309/534) and aligned well with previous reports of toward and away selectivity in MT.[13,14] We additionally quantified the strength of toward and away selectivity using an asymmetry index (AI) that compared responses to toward versus away motion across coherences for each neuron and cue condition (STAR Methods). The AI was previously used to quantify 3D motion selectivity in MT,[14] with positive values indicating a toward preference and negative values indicating an away preference. Based on the combined-cue AIs, a larger proportion of tuned neurons preferred toward rather than away motion (MT, 77%, n = 149/194; FST, 69%, n = 213/309; chi-squared test, both $p < 2.82 \times 10^{-11}$). This tendency was also observed in most of the isolated cue conditions (Figure S2) and may reflect the importance of detecting objects moving directly toward the head (see discussion). All subsequent analyses were performed using neurons with significant tuning in the combined-cue condition to test if that tuning reflected genuine 3D motion selectivity or lower-level 2D retinal-motion selectivity.

Although a large proportion of neurons showed significant selectivity in the combined-cue condition, such responses are insufficient to distinguish 3D motion selectivity from 2D retinal-motion selectivity (Figures 1D and 1E). Instead, comparisons must be made across cue conditions. Example responses to each of the four cue conditions are shown for four MT neurons in Figures 3A–3D and four FST neurons in Figures 3E–3H. Following the 2D retinal-motion predictions, most MT neurons (Figures 3A–3C) had: (1) negatively correlated left- and right-eye perspective cue responses and (2) combined-cue and stereoscopic cue responses that followed the dominant eye's perspective cue responses. A few MT neurons had responses that followed the 3D motion predictions (Figure 3D). In contrast, many FST neurons (Figures 3E–3G) had cue-invariant direction preferences, as predicted for 3D motion selectivity. The responses of other FST neurons followed the 2D motion predictions (Figure 3H). To test if 2D retinal-motion signals are transformed into a representation of 3D motion across an MT-to-FST hierarchy, the next section quantifies the prevalence of 2D and 3D selectivity in each area.

## Prevalent 3D-motion selectivity in FST but not MT

To determine the prevalence of 2D and 3D motion selectivity in MT and FST, we classified the neurons based on their responses to left- and right-eye perspective cues. For this analysis, neurons were required to show statistically significant tuning in both perspective cue conditions (MT, 55%, n = 106/194; FST, 51%, n = 157/309). To classify the neurons, we computed Z-transformed partial correlation coefficients[23,24] based on the predictions outlined in Figures 1D and 1E (STAR Methods). For each neuron, we first computed the correlation between the left- and right-eye perspective cue tuning curves aligned relative to the direction of 3D motion (as in Figure 3). This 3D motion correlation is positive for 3D selective neurons and negative for 2D

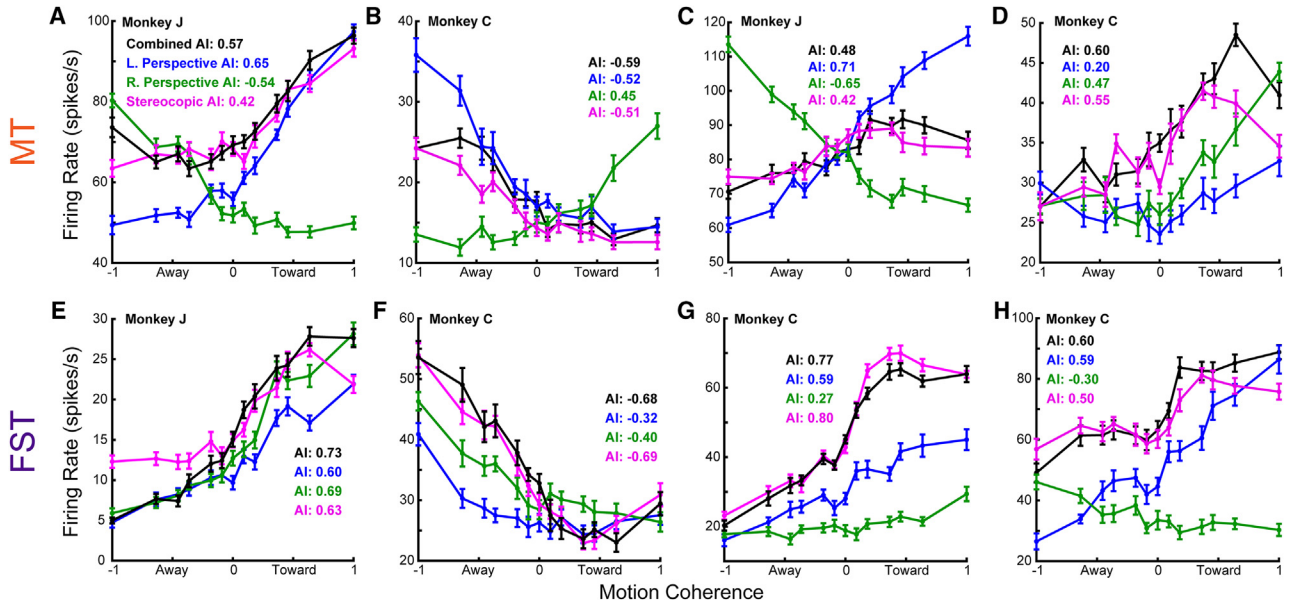



**Figure 3. Example MT and FST responses to 3D motion**

(A–C) Example MT neurons whose responses imply 2D retinal-motion selectivity (cf., Figure 1D). In (C), note that weak ocular dominance was associated with attenuated combined-cue and stereoscopic cue peak responses, consistent with motion opponency. Asymmetry index (AI) values for each cue condition are listed as insets.

(D) A rare MT neuron whose responses were consistent with 3D selectivity (cf., Figure 1E).

(E–G) Example FST neurons whose responses imply 3D motion selectivity.

(H) An FST neuron whose responses imply 2D retinal-motion selectivity. See also Figure S2.

selective neurons. We then aligned the tuning curves relative to the net direction of 2D retinal motion (i.e., in 2D retinal coordinates) by reflecting one of them about the 0% motion coherence and computed a 2D retinal-motion correlation. This correlation is negative for 3D selective neurons and positive for 2D selective neurons. Last, we Z-transformed the 2D ($Z_{2D}$) and 3D ($Z_{3D}$) motion correlations, allowing for the statistical classification of 2D versus 3D motion selectivity (Figure 4).

A majority of MT neurons were statistically classified as 2D retinal-motion selective (n = 80/106; 75%), while a minority were classified as 3D selective (n = 9/106; 8%). The predominance of 2D selectivity in MT was further evident in the distribution of classifications (Figure 4A, diagonal histogram), which had no indication of having multiple modes (Hartigan's dip test, p = 0.93). In stark contrast, the proportions of FST neurons that were classified as 2D (n = 66/157; 42%) or 3D (n = 58/157; 37%) selective were quite similar. Moreover, the distribution of classifications was bimodal (Hartigan's dip test, $p = 4.6 \times 10^{-3}$), consistent with separate subpopulations of 2D and 3D selective neurons. This possibility was further supported by distinct time courses of $Z_{3D}$–$Z_{2D}$ (Figure S3). After controlling for area, we found that 2D and 3D selective neurons had similar visual response latencies (linear mixed-effects model predicting latency from motion selectivity classification, area, and their interaction with monkey as a random effect; classification main effect, p = 0.81; area main effect, $p = 7.68 \times 10^{-8}$; interaction, p = 0.49). At least within FST (note the limited statistical power for MT), this may suggest that 2D motion and 3D motion are processed in parallel. These analyses revealed a hierarchical transformation from 2D to 3D motion selectivity between MT and FST that is reminiscent of the transformation of component to pattern motion selectivity between V1 and MT.[24]

Our predictions further included that 2D retinal-motion selective neurons should have responses to binocular (combined-cue and stereoscopic cue) 3D motion stimuli that positively associate with the dominant eye's perspective cue responses and negatively associate with those of the non-dominant eye. Importantly, the strength of OD is expected to moderate this relationship, such that it will be strongest for neurons with strong OD and weakest for neurons with weak OD, since neither eye can be reliably defined as dominant. To test these predictions, we quantified OD using a monocularity index (STAR Methods). We then used logistic regression to determine if the combined-cue and stereoscopic cue direction preferences (toward or away) of 2D selective neurons were predicted by the dominant (or non-dominant) eye perspective cue AI, log-transformed OD magnitude, and their interaction (Figure S4, top row). As predicted, the combined-cue preferences were positively associated with the dominant eye perspective cue AIs (MT, $b = 8.05$, $p = 1.7 \times 10^{-4}$; FST, $b = 9.68$, $p = 2.6 \times 10^{-3}$), and the magnitude of the effect increased with stronger OD in both MT (interaction, $b = 1.43$, $p = 8.3 \times 10^{-3}$) and FST (interaction, $b = 2.43$, $p = 2.6 \times 10^{-3}$). Importantly, the combined-cue preferences were negatively associated with the non-dominant eye perspective cue AIs (MT, $b = -8.67$, $p = 5.9 \times 10^{-5}$; FST, $b = -12.96$, $p = 1.3 \times 10^{-4}$), and the magnitude of the effect again increased with stronger OD in both MT (interaction, $b = -1.69$, $p = 2.4 \times 10^{-3}$) and FST (interaction, $b = -3.41$, $p = 5.3 \times 10^{-4}$). For the

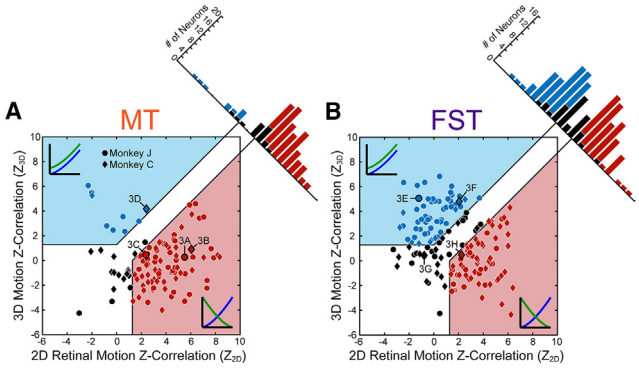



**Figure 4. Classification of 2D and 3D motion selectivity based on responses to eye-specific perspective cues**

(A) Classification of motion selectivity in MT. Statistical boundaries demarcate regions defining neurons as 2D retinal-motion selective (red region and points), 3D motion selective (blue region and points), or unclassified (white region with black points). Most MT neurons were 2D retinal-motion selective. The diagonal histogram shows the distribution of classifications.

(B) Same as (A), but for FST. Approximately equal proportions of FST neurons were selective for 2D and 3D motion. Note the bimodal distribution of classifications. See also Figures S3 and S4.

stereoscopic cue stimuli, the MT results were similar, such that the stereoscopic cue preferences were positively associated with the dominant eye perspective cue AIs ($b$ = 8.3, p = 3.3 × $10^{-3}$; interaction, $b$ = 1.34, p = 0.06) and negatively associated with the non-dominant eye perspective cue AIs ($b$ = −9.43, p = 1.6 × $10^{-3}$; interaction, $b$ = −1.61, p = 0.028). For FST, the stereoscopic cue results followed similar trends but were not significant for the dominant eye ($b$ = 2.55, p = 0.26; interaction, $b$ = 0.21, p = 0.80) or non-dominant eye ($b$ = −2.66, p = 0.26; interaction, $b$ = −0.22, p = 0.80), further distinguishing the two areas.

Last, for 3D selective neurons in both areas, the combined-cue and stereoscopic cue AIs were positively associated with both eyes' perspective cue responses, as predicted (Figure S4, bottom row). These results collectively imply that MT is largely specialized for processing 2D retinal motion, whereas FST contains a prominent representation of 3D motion.

## No systematic relationship between 3D motion and saccade direction selectivity

Having found greater prevalence of 3D motion selectivity as well as saccade direction selectivity in FST than in MT, we considered if there was a connection between the two properties. We first observed that relatively few neurons with combined-cue tuning also had saccade direction selectivity (MT, n = 21/194; FST, n = 58/309). Second, for neurons with both combined-cue tuning and saccade direction selectivity, the preferences were congruent (e.g., toward motion and downward saccades) with roughly the same frequency as they were opposite (MT, n = 9/21; FST, n = 37/58 were congruent). Third, neurons with congruent preferences were rarely classified as 3D selective (MT, n = 0/9; FST, n = 5/37). Last, for all neurons with saccade direction selectivity (MT, n = 35; FST, n = 98), we recomputed the AI for each cue condition using a shorter analysis window (750 ms) to remove potential saccade signals (Figure S1). The recomputed AIs were nearly identical to those from the full 1 s window (Pearson correlation; MT, r ≥ 0.98; FST, r ≥ 0.97 across all four cue conditions). These results collectively indicate that there was no systematic relationship between 3D motion and saccade direction selectivity.

## Robust linear decoding of 3D motion from FST, but not MT

We next assessed whether the hierarchical transformation of visual motion signals between MT and FST served to simplify the decoding of 3D motion. Intuitively, the conjunction of select combinations of MT neuronal responses could be used to infer 3D motion. For example, toward motion can be inferred from the simultaneous activation of a leftward selective MT neuron with right OD and a rightward selective MT neuron with left OD. However, this would require the decoder to possess explicit knowledge of utrocular (eye of origin) information. Indeed, previous work proposed that MT is the site of 3D motion decoding through a non-canonical mechanism that incorporated 2D motion selectivity and OD.[16] However, the above findings suggest that hierarchical computations performed across MT and FST might generalize, and therefore simplify, the decoding of 3D motion.[25]

To test this possibility and highlight the challenges of decoding 3D motion from MT, we trained linear decoders (Fisher's linear discriminant) to classify toward and away motion. Separate decoders were trained for MT and FST using the 2D and 3D motion selective subpopulations. To relate our findings to previous work, which emphasized the role of OD in 3D-motion processing,[10,16,17,26] each neuron's OD (monocularity index) was used to reassign the perspective cue conditions as dominant and non-dominant eye perspective cues. Training was performed using the combined-cue data and all motion coherences (STAR Methods), and we evaluated the performance of the decoders across cue conditions and motion coherences.

We first examined the MT results. For both the 2D (Figure 5A, top row) and the 3D (Figure 5A, bottom row) motion selective neurons, the performance of the decoders was consistently above chance and generally increased with motion coherence for the combined- (black curve), stereoscopic- (magenta curve), and dominant-eye-perspective- (yellow curve) cues. However, it is essential to emphasize that this pattern would be expected regardless of whether the neurons encoded 2D or 3D motion. The critical results to consider are instead how the decoders performed with the non-dominant eye perspective cues. For the 2D selective MT neurons, this performance was consistently below chance and became systematically worse with increasing motion coherence (cyan curve; Figure 5A, top row). This systematic confusion of toward and away motion indicates that the decoder extracted the net direction of 2D retinal motion. Even for the 3D selective MT neurons, performance was near chance for the non-dominant eye perspective cues (cyan curve; Figure 5A, bottom row). These results imply that if 3D motion was decoded from MT, then the decoding weights would need to dynamically change to reflect whether the left eye, right eye, or both eyes happened to be open; otherwise, opposite directions of 3D motion would be confused.

We next tested if FST could support generalized, cue-invariant decoding of 3D motion based on a fixed set of decoding weights.

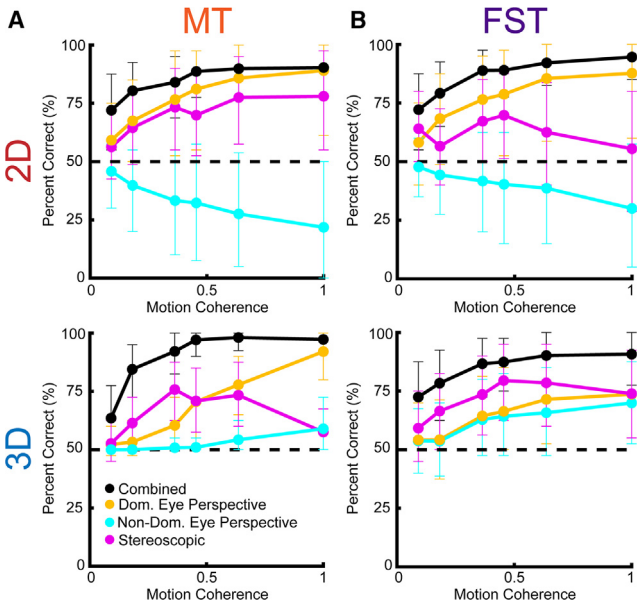



**Figure 5. Linear decoding of 3D motion direction from MT and FST**

(A) Decoding performance with 2D (top) and 3D (bottom) selective MT neurons as a function of motion coherence. Decoders were trained on the combined-cue responses and tested on the combined-cue (black), stereoscopic cue (magenta), dominant eye perspective cue (yellow), and non-dominant eye perspective cue (cyan) responses. Error bars are 95% confidence intervals.

(B) Same as (A), but for FST. Generalized, cue-invariant decoding of 3D motion was achieved only with 3D motion selective FST neurons. See also Figure S5.

Not surprisingly, the decoding of toward versus away motion from 2D selective FST neurons was qualitatively similar to that in MT. In particular, the direction of motion tended to be misclassified for the non-dominant eye perspective cues (Figure 5B, top row). However, when decoding from the 3D selective FST neurons, performance was consistently above chance and generally increased with coherence in each condition (Figure 5B, bottom row). Most critically, performance was highly similar for the dominant and non-dominant eye perspective cues, despite the stimuli having equal and opposite net retinal motions. The finding that cue-invariant decoding was most readily achieved with 3D selective FST neurons was further highlighted by decoding results using the full neuronal populations (Figure S5). These results thus substantiate that a hierarchical transformation of motion signals occurs across MT and FST and reveal that the transformation generalizes decoding by allowing 3D motion to be extracted using a single set of weights, regardless of the defining visual cues or which eye was open.

Last, reductions in decoding performance were apparent in the stereoscopic cue condition at high coherences for the 3D selective MT and both FST subpopulations. Similar effects were observed in single-neuron responses (Figure 3) and neurometric analyses presented in the next section. These observations may collectively reflect that FST neurons are often sensitive to more complex retinal velocity patterns than 2D translations[27] and that the 3D selective neurons in particular were sensitive to patterns of optic flow that naturally occur with 3D motion but which were absent in the stereoscopic cue condition (see discussion).

## 3D selective neurons account for behavioral sensitivity to 3D motion

The experiments additionally allowed for the direct comparison of behavioral and neuronal sensitivity to 3D motion. To this end, we quantified the monkeys' behavioral sensitivity for each cue condition by fitting psychometric functions to the proportion of toward reports at each motion coherence.[8] We similarly quantified how well an ideal observer could discriminate toward from away motion using single-neuron responses by calculating neurometric sensitivities for each condition (STAR Methods).[28,29] Using these measures, we compared the behavioral sensitivities to the neurometric sensitivities of 2Dand 3D selective neurons in MT and FST.

Median psychometric functions across recording sessions and neurometric functions across neurons are shown for each cue condition in Figure 6A. As expected from previous behavioral studies,[7,8] behavioral sensitivity in the combined-cue condition was greater than in any of the isolated cue conditions (linear mixed-effects model predicting log-transformed sensitivities with condition, eccentricity, and brain area as main effects and monkey as a random effect; $p = 1.3 \times 10^{-65}$). Across the isolated cue conditions, behavioral sensitivity to the two eyes' perspective cues was relatively similar (analogous linear mixed-effects model comparing left- and right-eye perspective cues; $p = 0.07$), and the monkeys were somewhat more sensitive to stereoscopic cues than perspective cues (analogous model comparing stereoscopic cues with both perspective cues; $p = 5.3 \times 10^{-5}$). There was no significant effect of area ($p \geq 0.48$ for all models), indicating that behavioral performance was similar across the MT and FST recording sessions. These findings reflect the known perceptual integration of perspective and stereoscopic cues to 3D motion.

The pattern of behavioral sensitivity across the four cue conditions was markedly different from the neurometric sensitivities of 2D selective MT and FST neurons (Figure 6A, cf., column 1 to columns 2 and 3). Most notably, neurometric sensitivities in the perspective cue conditions (i.e., during monocular viewing) exceeded those in the combined-cue and stereoscopic cue conditions (i.e., during binocular viewing; $p \leq 2.2 \times 10^{-13}$; linear mixed-effects models predicting log-transformed sensitivities with condition as a main effect and monkey as a random effect). Indeed, on individual sessions, the behavioral and neurometric sensitivities were typically negatively correlated across the four cue conditions (MT, mean = −0.33; FST, mean = −0.14; both means significantly less than 0, t test, $p \leq 0.021$). Thus, the sensitivities of 2D selective neurons did not account for the behavioral performance of the monkeys (Figure 6B, first two columns). Instead, the lower neurometric sensitivities in the binocular viewing conditions were consistent with 2D motion opponency[21,30–32] that reduced the overall response modulation when the two eyes saw opposite net 2D directions (e.g., see Figure 3C).

We next compared the behavioral sensitivities to the neurometric sensitivities of the 3D selective neurons. Similar to the monkeys, 3D selective neurons were most sensitive in the combined-cue condition (linear mixed-effects models predicting log-transformed sensitivities with condition as a main effect and monkey as a random effect; MT, $p = 0.035$; FST, $p = 3.9 \times 10^{-7}$). Neurometric sensitivities in the two perspective cue

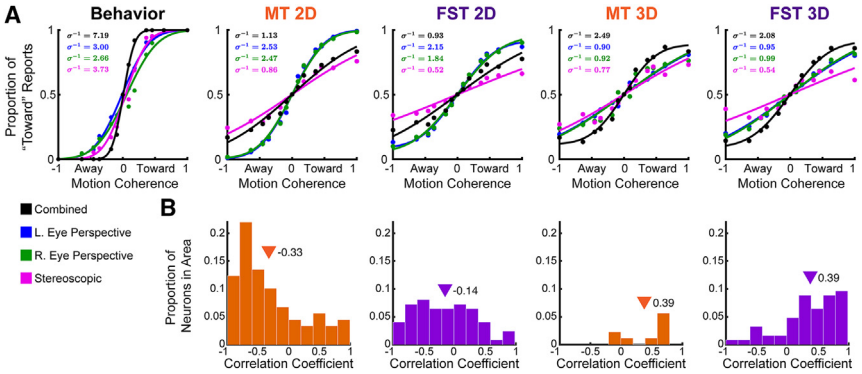



**Figure 6. Behavioral and neuronal sensitivity to 3D motion**

(A) Median psychometric and neurometric curves for each area (MT and FST) and subpopulation (2D and 3D).

(B) Distributions of Pearson correlation coefficients comparing simultaneously measured behavioral and neurometric sensitivities for each cue condition (arranged in columns following A).

conditions were similar (analogous model comparing left- and right-eye perspective cues; MT, $p = 0.67$; FST, $p = 0.83$). Sensitivities to the perspective and stereoscopic cues were also similar (analogous model comparing stereoscopic cues with both perspective cues; MT, $p = 0.95$; FST, $p = 0.42$) (Figure 6A, columns 4 and 5). On individual sessions, the behavioral and neurometric sensitivities were typically positively correlated across the four cue conditions (MT, mean = 0.39; FST, mean = 0.39; both means significantly greater than 0, t test, $p \leq 0.011$) (Figure 6B, last two columns). In the stereoscopic cue condition, 3D selective FST neurons showed reductions in discriminability at the highest coherences, consistent with the decoding results (Figure 5). While a similar effect was not behaviorally apparent, note that the behavior saturated at comparatively lower coherences (Figure 6A, first column), consistent with a ceiling effect in the relationship between neuronal response amplitude and inferred motion direction.[33,34] Overall, these results indicate that the behavioral sensitivities were best explained by the responses of 3D selective neurons and most strongly associated with FST due to the cross-area difference in prevalence of 3D selectivity.

The pattern of neurometric sensitivities across the cue conditions further distinguished 2D and 3D selective neurons. For 2D selective neurons, larger sensitivities were associated with monocular over binocular viewing conditions, consistent with motion opponency (which is commonplace in V1 and MT[30,31] but perhaps less so in FST[21]). In contrast, for 3D selective neurons, larger sensitivities were associated with the presence of multiple motion cues (as opposed to monocular versus binocular viewing), which is consistent with the theoretical principles of neuronal cue integration[33,35] and the perceptual integration of 3D motion cues.[7,8] To further explore these observations, we ran linear mixed-effects models to test if the strength of 3D versus 2D selectivity on a continuum ($Z_{3D}$–$Z_{2D}$) was predictable from the log-transformed neurometric sensitivity in each cue condition, area, and their interaction. Consistent with 2D motion opponency, stronger 2D selectivity was associated with greater left- and right-eye perspective cue (i.e., monocular viewing) sensitivities in both areas (sensitivity main effects, $p \leq 3.5 \times 10^{-4}$; interactions, $p \geq 0.12$). The relationships were qualitatively different for the combined- and stereoscopic cue conditions. Stronger 3D selectivity was associated with greater combined-cue sensitivity in both areas and the effect was much larger in FST ($b = 1.28$) than in MT ($b = 0.12$) (interaction, $p = 0.02$), consistent with FST containing a more balanced representation of 2D and 3D motion than MT, where neurons were predominantly 2D selective. Last, there was no significant association between $Z_{3D}$–$Z_{2D}$ and stereoscopic cue sensitivity in either area (sensitivity main effects and interactions, $p \geq 0.10$). This may reflect that the two measures quantify sensitivity to qualitatively different 3D motion cues and further suggests that processing at the level of 2D selective neurons fundamentally limits sensitivity to stereoscopic cues to 3D motion.

## Hierarchical model of 3D motion computation

The above findings revealed a hierarchical transformation between MT and FST that generalized the decoding of 3D motion. A potential mechanism for computing 3D motion is to integrate the output of 2D retinal-motion selective neurons with different direction preferences and ODs.[26,36] In that case, interocular 2D direction preference differences should be larger for 3D than for 2D selective neurons. To test this possibility, we compared the 2D direction preferences of individual neurons measured separately for the left and right eyes ($|\Delta 2D|$). As predicted for this mechanism, interocular differences in 2D direction preferences were significantly larger for 3D (MT, median = 39.81°, n = 9; FST, median = 37.25°, n = 58) than for 2D (MT, median = 6.36°, n = 80; FST, median = 6.50°, n = 66) selective neurons (Wilcoxon rank sum, $p = 2.8 \times 10^{-10}$) (Figure S6). Consistent with the greater prevalence of 3D selectivity in FST than in MT, $|\Delta 2D|$ values were significantly larger in FST than in MT (FST, median = 15.67°, n = 157; MT, median = 7.14°, n = 106; Wilcoxon rank sum, $p = 2.6 \times 10^{-4}$). These results thus confirmed that larger interocular 2D direction preference differences were indeed associated with 3D motion selectivity.

Given these empirical results, we used computational modeling to explore if integrating the output of neurons with different 2D retinal-motion direction preferences and ODs could broadly account for 3D motion selectivity spanning the space of 3D motion directions (STAR Methods). The first layer simulated a population of 2D retinal-motion selective units that were either left- or right-eye dominant (equal proportions) and jointly tuned for 2D direction and speed. The incorporation of speed tuning was necessary to enable the computation of 3D motion trajectories whose two retinal projections differ in both direction and speed. The tuning properties of the units were identical for the two eyes, except for an amplitude difference because of OD. The second layer simulated a population of units that linearly integrated the output of pairs of first-layer, 2D selective units, followed by an expansive non-linearity (Figure 7A).

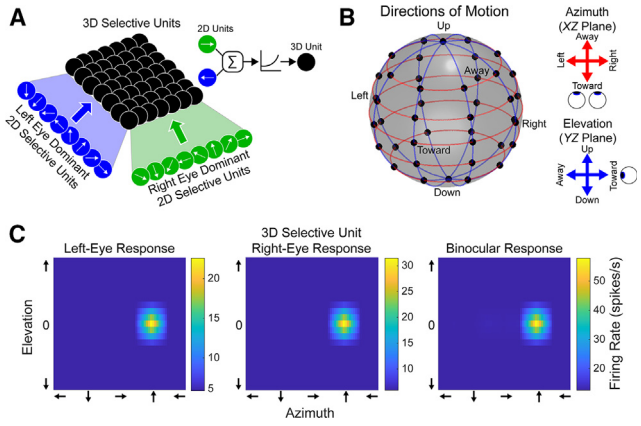



**Figure 7. Hierarchical model of 3D motion encoding**

(A) Model schematic. The first layer included two populations of 2D direction selective units with left- (blue) or right- (green) eye ocular dominance. Pairs of first-layer units with different 2D direction preferences and ODs were linearly combined followed by an expansive non-linearity (inset) in the second layer to create 3D selective units (black).

(B) Directions of motion. Directions were sampled across eight azimuths (red) and seven elevations (blue), for a total of 42 unique trajectories.

(C) Responses of a second-layer unit that preferred away motion, plotted using an equal-area Lambert projection. Yellow hues indicate higher firing rates and blue hues indicate lower firing rates. The preference for away motion was independent of whether the left eye (left image), the right eye (middle image), or both eyes (right image) were visually stimulated, and therefore the unit was cue invariant. See also Figures S6 and S7.

We then assessed if the model was sufficient to build a population of units whose preferences spanned the space of 3D directions (Figure 7B) and whether that selectivity was cue invariant. The 3D direction tuning curve of a second-layer unit tuned for away motion is shown in Figure 7C. Importantly, the 3D direction preference was the same regardless of whether the left, the right, or both eyes were visually stimulated. This was not unique to away (or toward) preferring units. Across the second layer, the preferences spanned the space of 3D directions and were cue invariant (Figure S7). Thus, the integration of 2D retinal-motion selective units with different ODs was sufficient to create a population of 3D selective units whose preferences spanned the space of 3D directions.

## DISCUSSION

Transforming ambiguous 2D retinal signals into behaviorally relevant representations of 3D motion is an essential function of the visual system.[4] While the hierarchical organization of MT and FST has long been anatomically known,[37,38] the functional significance of the pathway has remained challenging to discern.[21,27,39] Here we discovered that 2D retinal-motion signals are transformed across MT and FST into a robust representation of 3D motion. Importantly, we showed that the transformation generalizes the decoding of motion signals, such that a single decoder could extract the direction of 3D motion irrespective of the defining cues or which eye was open. Last, we showed how the transformation could be implemented through a linear-non-linear cascade.

Previous reports came to opposite conclusions regarding whether MT represents 3D[13,14] or 2D[12] motion. Multiple lines of evidence in the current study converged to indicate that 3D motion selectivity is rare in MT. First, for the vast majority of MT neurons, the apparent 3D direction preference was opposite for left- and right-eye perspective cue stimuli, indicating that the responses were driven by the net direction of 2D retinal motion. Second, responses to 3D motion in the binocular viewing conditions were well explained by lower-level properties of 2D direction selectivity and OD. Third, our results showed that decoding 3D motion from MT would require highly dynamic utrocular (eye of origin) information, but monkeys do not use such information to discriminate 3D motion.[8] In the absence of utrocular information, decoding 3D motion from MT would result in systematic confusions of toward and away motion, depending on whether one eye (or the other) was closed. While motion-in-depth confusions have been observed in human psychophysics,[40–42] they are distinct from the current findings. Specifically, confusions in decoding 3D motion from MT were associated with non-dominant eye perspective cues, whereas confusions in human perceptual data occurred in low-contrast binocular viewing conditions. Fourth, the neurometric sensitivities of MT neurons were consistently larger in the monocular viewing conditions than in the binocular viewing conditions. This finding is consistent with 2D retinal-motion opponency, in which opposing dichoptic motion signals during the binocular presentation of toward/away motion suppressed neuronal responses. As such, the responses of MT neurons did not account for the behavioral sensitivity of the monkeys, which was greater in the combined-cue condition than the isolated cue conditions.

Previous work suggested that an interaction between 2D direction selectivity and OD supports non-canonical encoding/decoding of 3D motion in MT.[16] Although these physiological properties are consistent with our results, the model would be highly inefficient, since the decoder would not generalize across visual cues or viewing conditions. The model also has untenable implications. In particular, if the mechanism was sufficient for defining 3D motion selectivity, then it would also be necessary to classify 2D motion selective retinal ganglion cells as 3D selective, since they too will selectively respond to toward or away motion as a simple consequence of projective geometry. Thus, a more parsimonious conclusion is that MT is largely specialized for 2D retinal-motion processing[12,15,43,44] and that the persistence of OD in MT enables downstream computations that support robust 3D motion processing.

Our finding of 3D motion selectivity in FST is consistent with a recent neuroimaging study that suggested that neurons in the area encode stereomotion.[15] Interestingly, we found roughly equal proportions of 2D and 3D motion selective neurons in FST. This finding is consistent with the area containing a broad representation of 3D motion, since 2D directions are a subset of the larger space of 3D directions. Moreover, the behavioral sensitivity of the monkeys across the four cue conditions was best explained by the neurometric sensitivities of 3D selective FST neurons—the same neurons whose responses could be linearly decoded to provide a generalized, cue-invariant readout of 3D motion. Indeed, only the sensitivity of the 3D selective neurons reflected the number of visual cues signaling 3D motion,

implying cue integration consistent with the animal's behavior. This starkly contrasts with the 2D selective neurons, whose sensitivity reflected whether the viewing conditions were monocular (more sensitive) or binocular (less sensitive), consistent with 2D motion opponency. These findings collectively suggest that FST is a prominent site of 3D motion processing, which may support the integration of 3D motion cues in a statistically optimal fashion.[8]

How might the 2D-to-3D transformation of visual motion signals be accomplished across MT and FST? Theoretical work suggests that 3D motion selectivity may arise, in part, through interocular differences in 2D direction preferences.[26,36] In line with this, FST neurons showed larger differences in interocular 2D direction preferences than MT neurons. Although the experiments only tested 3D motion trajectories that intersected the cyclopean eye, our simulations showed that a mechanism based on interocular differences in 2D direction and speed tuning could support a population of units representing the full span of 3D directions. It will be interesting for future work to characterize the distribution of 3D directions represented by FST and relate that distribution to the natural scene statistics of 3D motion. However, doing so will also require the natural scene statistics of 3D motion to be characterized. The current findings suggest that motions toward the monkey will be overrepresented, which may imply that the distribution will reflect a combination of scene statistics and the behavioral relevance of specific trajectories. Interestingly, we found that, although MT was largely specialized for 2D motion processing, neurons with significant combined-cue tuning tended to have 2D direction preferences and ODs consistent with toward motion (rightward motion + left OD, n = 67; leftward motion + right OD, n = 57) as opposed to away motion (leftward motion and left OD, n = 33; rightward motion + right OD, n = 37). This may have contributed to the finding that more downstream 3D selective FST neurons preferred toward rather than away motion.

Previous computational work demonstrated that 3D motion selectivity could be achieved by integrating the output of units with opposite horizontal direction preferences and ODs.[16,17,26,36] We extended that work to show that a population of units whose direction preferences spanned the space of 3D motion could be created using a cascade model in which 2D motion selective units with OD were selectively integrated. This was achieved by allowing the integrated units to have different 2D direction (not just opposite) preferences, speed preferences, and ODs. Importantly, we found that these properties were sufficient to generalize the creation of 3D direction selectivity. The model did not explore, however, if this selectivity was also shaped by computations such as motion opponency, earlier stages of binocular integration, or divisive normalization. These computations are widely implicated in the hierarchical processing of retinal motion as well as motion in depth.[17,45,46] They may also be important to build selectivity for eye-specific optic flow patterns (see limitations of the study), which may speculatively lead to reduced responses to high-coherence stereoscopic cue stimuli that lack natural perspective cues. Exploring these computations in the context of 3D motion processing is therefore likely to further advance our understanding of how motion signals are transformed across the V1-MT-FST hierarchy.

Although little is currently known about the role of FST in visual processing, it has been linked to object recognition and segmentation, attention, change detection, and structure-from-motion processing.[15,17,20,21,27,39,47,48] Taking this together with the present findings and evidence of connections to visual areas in the ventral pathway,[38,49,50] it is conceivable that FST contributes to the binding of object identity and 3D motion, enabling the tracking of specific objects through space.[4] Interestingly, inactivation of the superior colliculus (SC) was recently shown to modulate FST's activity and selectivity.[39] In line with this, we observed saccade-related activity in FST that preceded eye movements in an overlap saccade task. Future work can determine whether that activity is associated with FST's connection to the SC and ascertain if it is functionally related to selective spatial attention.

Several visual areas, in addition to FST, are known to contain 3D motion selective neurons.[51–53] Some of these areas, such as MSTd, are strongly implicated in self-motion processing. Compared with the large RFs of neurons in areas that support self-motion processing, neurons in the lateral/ventral portion of MST (MSTl/v) and FST have smaller RFs that may be more suitable for processing 3D object motion.[4,54–57] This possibility is consistent with our finding that small (3° diameter) stimuli elicited robust responses from FST neurons. The functional relationship between FST and MSTl/v in object-motion processing and the potential contributions of FST to self-motion processing are important topics for future studies.

### Limitations of the study

This study extends our understanding of the cortical areas and hierarchical computations that contribute to visual motion processing.[4] Previous work revealed hierarchical processing of 2D retinal motion, in which component motion signals extracted at the level of V1 were integrated to form pattern motion representations in MT.[24,58] The current findings reveal a subsequent stage of motion integration at the level of FST that transforms 2D retinal motion into a behaviorally relevant, cue-invariant representation of 3D motion. While the current study demonstrated a functional association between FST and 3D motion processing, it did not test if FST causally contributes to 3D perception. The experiments also only assessed selectivity for motion toward and away from the cyclopean eye. Knowledge of the shape and bandwidth of FST tuning curves over the span of 3D directions can further constrain models that capture the contributions of additional computations and areas to 3D motion processing. Such information would also be valuable for investigating the coordinate frame(s) in which FST neurons represent 3D motion. Last, an intriguing possibility to explore is that selectivity for eye-specific perspective cues to 3D motion (optic flow) is achieved computationally through non-linear RF structures that are distinct for each eye.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal preparation
- METHOD DETAILS
  - Behavioral control and stimulus presentation
  - Visual stimuli
  - Receptive field mapping
  - Behavioral tasks
  - Neuronal recordings
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral data analysis
  - Visual response latency
  - Saccade analysis
  - Asymmetry Index (AI)
  - Monocularity Index
  - 2D motion analysis
  - Classification of 2D retinal-motion and 3D motion selectivity
  - Linear population decoding
  - Neurometric sensitivity
  - Hierarchical model of 3D motion encoding
  - Vergence control

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113524.

## ACKNOWLEDGMENTS

We thank Zikang Zhu for assistance with monkey training. The work was supported by National Institutes of Health grants EY035005 and EY029438. L.W.T. was supported by National Science Foundation LUCID Training Program 1545481, McPherson Eye Research Institute Graduate Student Support Initiative and Expanding Our Vision 2020 awards, and National Institutes of Health grant NS105602. B.K. was supported by National Institutes of Health grant NS128586. B.R. was supported by ASPIRE, the technology program management pillar of Abu Dhabi's Advanced Technology Research Council, via the ASPIRE Precision Medicine Research Institute Abu Dhabi (VRI-20-10) and the NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute grant CG012. Further support was provided by National Institutes of Health grant P51OD011106 to the Wisconsin National Primate Research Center.

## AUTHOR CONTRIBUTIONS

Conceptualization, L.W.T., B.R., and A.R.; methodology, L.W.T., B.K., B.R., and A.R.; investigation, L.W.T.; formal analysis, L.W.T., B.K., and A.R.; data curation, L.W.T. and B.K.; writing – original draft, L.W.T., B.R., and A.R.; writing – review and editing, L.W.T., B.K., B.R., and A.R.; visualization, L.W.T.; supervision, B.R. and A.R.; resources, A.R.; funding acquisition, A.R.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 23, 2023
Revised: October 25, 2023
Accepted: November 15, 2023

## REFERENCES

1. Barton, R.A. (2004). Binocularity and brain evolution in primates. Proc. Natl. Acad. Sci. USA *101*, 10113–10115.
2. Bradley, D.C., and Goyal, M.S. (2008). Velocity computation in the primate visual system. Nat. Rev. Neurosci. *9*, 686–695.
3. Cormack, L.K., Czuba, T.B., Knöll, J., and Huk, A.C. (2017). Binocular mechanisms of 3D motion processing. Annu. Rev. Vis. Sci. *3*, 297–318.
4. Rosenberg, A., Thompson, L.W., Doudlah, R., and Chang, T.Y. (2023). Neuronal representations supporting three-dimensional vision in nonhuman primates. Annu. Rev. Vis. Sci. *9*, 337–359.
5. Czuba, T.B., Rokers, B., Guillet, K., Huk, A.C., and Cormack, L.K. (2011). Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion through depth. J. Vis. *11*, 18.
6. Thompson, L., Ji, M., Rokers, B., and Rosenberg, A. (2019). Contributions of binocular and monocular cues to motion-in-depth perception. J. Vis. *19*, 2.
7. Fulvio, J.M., Ji, M., Thompson, L., Rosenberg, A., and Rokers, B. (2020). Cue-dependent effects of VR experience on motion-in-depth sensitivity. PLoS One *15*, e0229929.
8. Thompson, L.W., Kim, B., Zhu, Z., Rokers, B., and Rosenberg, A. (2021). Perspective cues make eye-specific contributions to 3D motion perception. J. Cognit. Neurosci. *34*, 192–208.
9. Born, R.T., and Bradley, D.C. (2005). Structure and function of visual area MT. Annu. Rev. Neurosci. *28*, 157–189.
10. Zeki, S.M. (1974). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. J. Physiol. *242*, 827–841.
11. Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. J. Neurophysiol. *52*, 1106–1130.
12. Maunsell, J.H., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. J. Neurophysiol. *49*, 1148–1167.
13. Czuba, T.B., Huk, A.C., Cormack, L.K., and Kohn, A. (2014). Area MT encodes three-dimensional motion. J. Neurosci. *34*, 15522–15533.
14. Sanada, T.M., and DeAngelis, G.C. (2014). Neural representation of motion-in-depth in area MT. J. Neurosci. *34*, 15508–15521.
15. Héjja-Brichard, Y., Rima, S., Rapha, E., Durand, J.B., and Cottereau, B.R. (2020). Stereomotion processing in the nonhuman primate brain. Cerebr. Cortex *30*, 4528–4543.
16. Bonnen, K., Czuba, T.B., Whritner, J.A., Kohn, A., Huk, A.C., and Cormack, L.K. (2020). Binocular viewing geometry shapes the neural representation of the dynamic three-dimensional environment. Nat. Neurosci. *23*, 113–121.
17. Baker, P.M., and Bair, W. (2016). A model of binocular motion integration in MT neurons. J. Neurosci. *36*, 6563–6582.
18. Petro, L.S., Smith, F.W., Schyns, P.G., and Muckli, L. (2013). Decoding face categories in diagnostic subregions of primary visual cortex. Eur. J. Neurosci. *37*, 1130–1139.
19. Gattass, R., and Gross, C.G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. J. Neurophysiol. *46*, 621–638.
20. Desimone, R., and Ungerleider, L.G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol. *248*, 164–189.
21. Rosenberg, A., Wallisch, P., and Bradley, D.C. (2008). Responses to direction and transparent motion stimuli in area FST of the macaque. Vis. Neurosci. *25*, 187–195.
22. Kolster, H., Mandeville, J.B., Arsenault, J.T., Ekstrom, L.B., Wald, L.L., and Vanduffel, W. (2009). Visual field map clusters in macaque extrastriate visual cortex. J. Neurosci. *29*, 7031–7039.

23. Smith, M.A., Majaj, N.J., and Movshon, J.A. (2005). Dynamics of motion signaling by neurons in macaque area MT. Nat. Neurosci. *8*, 220–228.

24. Khawaja, F.A., Tsui, J.M.G., and Pack, C.C. (2009). Pattern motion selectivity of spiking outputs and local field potentials in macaque visual cortex. J. Neurosci. *29*, 13702–13709.

25. Ni, A.M., Huang, C., Doiron, B., and Cohen, M.R. (2022). A general decoding strategy explains the relationship between behavior and correlated variability. Elife *11*, e67258.

26. Sabatini, S.P., and Solari, F. (2004). Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. Neurocomputing *58–60*, 865–872.

27. Mysore, S.G., Vogels, R., Raiguel, S.E., Todd, J.T., and Orban, G.A. (2010). The selectivity of neurons in the macaque fundus of the superior temporal area for three-dimensional structure from motion. J. Neurosci. *30*, 15491–15508.

28. Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.

29. Elmore, L.C., Rosenberg, A., DeAngelis, G.C., and Angelaki, D.E. (2019). Choice-related activity during visual slant discrimination in macaque CIP but not V3A. eNeuro *6*. ENEURO.0248, 18.2019.

30. Snowden, R.J., Treue, S., Erickson, R.G., and Andersen, R.A. (1991). The response of area MT and V1 neurons to transparent motion. J. Neurosci. *11*, 2768–2785.

31. Qian, N., and Andersen, R.A. (1994). Transparent motion perception as detection of unbalanced motion signals. J. Neurosci. *14*, 7367–7380.

32. Qian, N., and Andersen, R.A. (1995). V1 responses to transparent and nontransparent motions. Exp. Brain Res. *103*, 41–50.

33. Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nat. Neurosci. *9*, 1432–1438.

34. Chang, T.Y., Doudlah, R., Kim, B., Sunkara, A., Thompson, L.W., Lowe, M.E., and Rosenberg, A. (2020). Functional links between sensory representations, choice activity, and sensorimotor associations in parietal cortex. Elife *9*, e57968.

35. Seilheimer, R.L., Rosenberg, A., and Angelaki, D.E. (2014). Models and processes of multisensory cue combination. Curr. Opin. Neurobiol. *25*, 38–46.

36. Poggio, G.F., and Talbot, W.H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. J. Physiol. *315*, 469–492.

37. Ungerleider, L.G., and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. J. Comp. Neurol. *248*, 190–222.

38. Boussaoud, D., Ungerleider, L.G., and Desimone, R. (1990). Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. J. Comp. Neurol. *296*, 462–495.

39. Bogadhi, A.R., Katz, L.N., Bollimunta, A., Leopold, D.A., and Krauzlis, R.J. (2021). Midbrain activity shapes high-level visual properties in the primate temporal cortex. Neuron *109*, 690–699.e5.

40. Welchman, A.E., Tuck, V.L., and Harris, J.M. (2004). Human observers are biased in judging the angular approach of a projectile. Vis. Res. *44*, 2027–2042.

41. Fulvio, J.M., Rosen, M.L., and Rokers, B. (2015). Sensory uncertainty leads to systematic misperception of the direction of motion in depth. Atten. Percept. Psychophys. *77*, 1685–1696.

42. Rokers, B., Fulvio, J.M., Pillow, J.W., and Cooper, E.A. (2018). Systematic misperceptions of 3-D motion explained by Bayesian inference. J. Vis. *18*, 23.

43. Lagae, L., Maes, H., Raiguel, S., Xiao, D.K., and Orban, G.A. (1994). Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. J. Neurophysiol. *71*, 1597–1626.

44. Nakhla, N., Korkian, Y., Krause, M.R., and Pack, C.C. (2021). Neural selectivity for visual motion in macaque area V3A. eNeuro *8*. ENEURO.0383, 20.2020.

45. Rust, N.C., Mante, V., Simoncelli, E.P., and Movshon, J.A. (2006). How MT cells analyze the motion of visual patterns. Nat. Neurosci. *9*, 1421–1431.

46. Simoncelli, E.P., and Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vis. Res. *38*, 743–761.

47. Kilintari, M., Raos, V., and Savaki, H.E. (2014). Involvement of the superior temporal cortex in action execution and action observation. J. Neurosci. *34*, 8999–9011.

48. Bogadhi, A.R., Bollimunta, A., Leopold, D.A., and Krauzlis, R.J. (2019). Spatial Attention Deficits Are Causally Linked to an Area in Macaque Temporal Cortex. Curr. Biol. *29*, 726–736.e4.

49. Saleem, K.S., Suzuki, W., Tanaka, K., and Hashikawa, T. (2000). Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. J. Neurosci. *20*, 5083–5101.

50. Ungerleider, L.G., Galkin, T.W., Desimone, R., and Gattass, R. (2008). Cortical connections of area V4 in the macaque. Cerebr. Cortex *18*, 477–499.

51. Chen, A., Gu, Y., Liu, S., DeAngelis, G.C., and Angelaki, D.E. (2016). Evidence for a causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception. J. Neurosci. *36*, 3789–3798.

52. Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2016). Joint representation of translational and rotational components of optic flow in parietal cortex. Proc. Natl. Acad. Sci. USA *113*, 5077–5082.

53. Gu, Y., Deangelis, G.C., and Angelaki, D.E. (2012). Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. J. Neurosci. *32*, 2299–2313.

54. Komatsu, H., and Wurtz, R.H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. J. Neurophysiol. *60*, 580–603.

55. Tanaka, K., Sugita, Y., Moriya, M., and Saito, H. (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.

56. Eifuku, S., and Wurtz, R.H. (1998). Response to motion in extrastriate area MSTl: center-surround interactions. J. Neurophysiol. *80*, 282–296.

57. Sasaki, R., Angelaki, D.E., and DeAngelis, G.C. (2019). Processing of object motion and self-motion in the lateral subdivision of the medial superior temporal area in macaques. J. Neurophysiol. *121*, 1207–1221.

58. Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1985). The analysis of visual moving patterns. Pattern recognition mechanisms, 117–151.

59. Kim, B., Kenchappa, S.C., Sunkara, A., Chang, T.-Y., Thompson, L., Doudlah, R., and Rosenberg, A. (2019). Real-time experimental control using network-based parallel processing. Elife *8*, e40231.

60. Kleiner, M., Brainard, D.H., Pelli, D.G., Broussard, C., Wolf, T., and Niehorster, D. (2007). What's new in Psychtoolbox-3? Perception *36*, S14.

61. Rosenberg, A., Cowan, N.J., and Angelaki, D.E. (2013). The visual representation of 3D object orientation in parietal cortex. J. Neurosci. *33*, 19352–19361.

62. Rosenberg, A., and Angelaki, D.E. (2014). Gravity influences the visual representation of object tilt in parietal cortex. J. Neurosci. *34*, 14170–14180.

63. Rosenberg, A., and Angelaki, D.E. (2014). Reliability-dependent contributions of visual orientation cues in parietal cortex. Proc. Natl. Acad. Sci. USA *111*, 18043–18048.

64. Doudlah, R., Chang, T.Y., Thompson, L.W., Kim, B., Sunkara, A., and Rosenberg, A. (2022). Parallel processing, hierarchical transformations, and sensorimotor associations along the 'where' pathway. Elife *11*, e78712.

65. Cooper, E.A., van Ginkel, M., and Rokers, B. (2016). Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. J. Vis. *16*, 5.

66. Allen, B., Haun, A.M., Hanley, T., Green, C.S., and Rokers, B. (2015). Optimal combination of the binocular cues to 3D motion. Invest. Ophthalmol. Vis. Sci. *56*, 7589–7596.

67. Van Essen, D.C., Lewis, J.W., Drury, H.A., Hadjikhani, N., Tootell, R.B., Bakircioglu, M., and Miller, M.I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. Vis. Res. *41*, 1359–1378.

68. Gray, C.M., Maldonado, P.E., Wilson, M., and McNaughton, B. (1995). Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. J. Neurosci. Methods *63*, 43–54.

69. Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H., and Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J. Neurophysiol. *84*, 401–414.

70. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11.

71. Schütt, H.H., Harmeling, S., Macke, J.H., and Wichmann, F.A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vis. Res. *122*, 105–123.

72. Tailby, C., Majaj, N.J., and Movshon, J.A. (2010). Binocular integration of pattern motion signals by MT neurons and by human observers. J. Neurosci. *30*, 7344–7349.

73. Sasaki, R., Anzai, A., Angelaki, D.E., and DeAngelis, G.C. (2020). Flexible coding of object motion in multiple reference frames by parietal cortex neurons. Nat. Neurosci. *23*, 1004–1015.

74. Nover, H., Anderson, C.H., and DeAngelis, G.C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J. Neurosci. *25*, 10049–10060.

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

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Experimental models: Organisms/strains | | |
| Rhesus monkeys (Macaca mulatta) | Wisconsin National Primate Research Center | https://primate.wisc.edu |
| Deposited data | | |
| Neuronal data | This study | OSF: https://osf.io/qvts8/ |
| Software and algorithms | | |
| MATLAB | Mathworks | RRID: SCR_001622 |
| Plexon | Plexon, Inc. | RRID: SCR_000012 |
| REC-GUI | Kim et al.[59] | RRID: SCR_019008 |
| Psychtoolbox | Kleiner et al.[60] | RRID: SCR_002881 |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by Ari Rosenberg (ari.rosenberg@wisc.edu).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- Supporting neuronal data have been made available in a public repository (https://osf.io/qvts8/) as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Animal preparation

Two male rhesus macaque monkeys (*Macaca mulatta*; Monkey J: 6.2 years of age, 8.4 kg in weight; Monkey C: 5.8 years, 9.0 kg) were used in this study. All surgeries and experimental procedures were approved by the Institutional Animal Care and Use Committee at the University of Wisconsin–Madison (Protocol G005229) and were in accordance with NIH guidelines. A head-stabilizing Delrin ring was surgically attached to the skull under general anesthesia.[29,34,61–64] After recovery, the monkeys were trained to sit in a primate chair and fixate visual targets within 2° version and 1° vergence windows for liquid reward. Both monkeys previously participated in a study which used the same stimuli and discrimination task to assess behavioral sensitivity to perspective and stereoscopic cues to 3D motion across the visual field.[8]

### METHOD DETAILS

#### Behavioral control and stimulus presentation

Experimental control was performed using the REC-GUI software (RRID:SCR_019008).[59] Stimuli were rear projected onto a polarization preserving screen (Screen-Tech e.K.) using a DLP LED projector (PROPixx; VPixx Technologies, Inc.) with a resolution of 1920 x 1080 pixels (58° x 35°) at 120 Hz. The screen was 57 cm from the monkey. Polarized glasses were worn. A phototransistor circuit was used to confirm the synchronization of the left- and right-eye images, and to align neuronal responses to the stimulus onset. The stimuli were created in MATLAB R2015a using Psychtoolbox 3[60] and rendered with anti-aliasing using an NVIDIA 1080T graphics card on a Linux workstation. Eye tracking was performed at 1,000 Hz (EyeLink 1000 plus, SR Research).

#### Visual stimuli

The 3D motion stimuli were previously described in detail.[8] Briefly, they depicted a volume with dots moving toward or away from the cyclopean eye. Twenty-two dots, 11 light (113 cd/m$^2$) and 11 dark (0.04 cd/m$^2$), were initialized with pseudorandom $x$, $y$, and $z$ positions within the volume. The volume was oriented toward the cyclopean eye, spanned ±1° of horizontal disparity (~42.5 cm of depth), and corresponded to a 3° diameter circular aperture on the screen. The background was gray (60 cd/m$^2$). Seven motion

coherences were used: 0, 0.09, 0.18, 0.36, 0.45, 0.64, and 1 (0/22, 2/22, 4/22, 8/22, 10/22, 14/22, and 22/22 signal/total dots, respectively). Signal and noise dots were pseudorandomly assigned each stereoscopic frame-pair. Signal dots moved toward or away from the monkey with an average retinal speed of ~4.2°/s while noise dots were reassigned to pseudorandom positions within the volume. The signal dot speed approximately maximized perceptual sensitivity to 3D motion[8,65] (Figure S8).

Four 3D motion cue conditions were included: (1) combined-cue stimuli which contained perspective and stereoscopic cues, (2) left-eye perspective cue stimuli, (3) right-eye perspective cue stimuli, and (4) stereoscopic cue stimuli. The perspective cues included optic flow and changes in the retinal density and size of the visual elements. Due to the horizontal offset of the eyes, the optic flow patterns for each eye had foci of expansion/contraction that were horizontally offset in opposite directions on the screen. Thus, the perspective cues were eye-specific, and left- and right-eye stimuli had opposite net horizontal motion directions. The stereoscopic cues included changing disparity and interocular velocity differences.[66] Dot disparities were uniformly distributed throughout the stimulus volume. It was therefore impossible to discriminate the direction of motion or account for neuronal selectivity to the stimuli based on the distribution of instantaneous disparities, dot sizes, or dot density on the screen. Stimulus examples are provided in an Extended Data Movie (https://osf.io/dsjza).

The 2D motion stimuli (22 dots, 11 light and 11 dark) were presented monocularly and binocularly within a 3° diameter circular aperture at eight directions (0° to 315° in 45° steps). The horizontal (0° and 180°) monocular conditions were equivalent to monocular views of the stereoscopic cue stimuli at 100% coherence. The 2D stimuli were presented at a speed of ~4.2°/s for all neurons, as well as at 12.6°/s for 165 MT and 523 FST neurons.

### Receptive field mapping

Initial receptive field (RF) estimates were obtained with hand-mapping techniques using random dot patches, drifting gratings, and/or oriented bars. Based on the estimates, a rectangular region with ~400-600 square patches was defined to cover and extend beyond the RF. Dark and light squares were flashed (150 ms duration) in an alternating sequence at pseudorandom locations during fixation of a target. A liquid reward was provided after every 20 flashes. If fixation was broken, the monkey was required to fixate the target for 300 ms before the sequence resumed. A single repetition was completed when light and dark squares had been flashed at each location. At least four blocks were completed each session. Responses at each grid location were averaged over blocks (regardless of brightness). A 2D Gaussian contour was fit to the average responses. The mean of the Gaussian was used as an estimate of the RF center. The sigma values defined principal axes demarcating an ellipse used to calculate the RF area.

### Behavioral tasks

We first measured 2D direction selectivity during passive fixation by pseudorandomly presenting eight directions of motion both monocularly and binocularly and at either one or two speeds. The monkey received a liquid reward after every third stimulus.

The 3D motion stimuli were then presented as part of a toward/away motion discrimination task. The trial structure proceeded as follows. First, a monkey acquired fixation on a red circular target (0.1°, 20.5 cd/m$^2$) at the center of the screen and aligned with the cyclopean eye. After 300 ms, a stimulus appeared for 1 s while fixation was maintained. The fixation target then disappeared and two red choice targets (0.1°, 20.5 cd/m$^2$) appeared 9° above/below the fixation target. The direction of motion was indicated as 'toward' or 'away' by making a saccade to the lower or upper target, respectively. A liquid reward was provided for correct responses. Responses to 0% coherence trials were rewarded pseudorandomly with 50% probability.

Interleaved with the 3D motion stimuli were presentations of monocular and binocular horizontal (0° and 180°) 2D motion, without the choice component of the trial. Also interleaved were trials of an overlap saccade task. On each trial of that task, one of two saccade targets was presented at locations corresponding to the choice targets in the motion discrimination task for 1 s while fixation was maintained on a central fixation target. The fixation target then disappeared, and the monkey made an eye movement to the saccade target for a liquid reward.

The inter-trial interval was 1.5 s. A trial was aborted without reward if fixation was broken before the end of the trial, or if a choice/saccade was not made within 500 ms. Stimuli were presented in a block structure, with completion of one trial of each condition required to finish the block. Aborted trials were reshuffled into the block until all were completed. The next block then began. A minimum of 5 blocks (mean = 6.2) of the 2D direction and 10 blocks (mean = 21.3) of the 3D motion discrimination/saccade tasks were required for inclusion in the dataset.

### Neuronal recordings

Areas MT and FST were identified based on anatomical and functional properties. The CARET software was used to register structural MRIs to the F99 atlas[67] and to estimate electrode trajectories. MT was distinguished from FST according to its well-established topography, where RFs are organized along a peripheral to foveal visual field representation moving posterior to anterior.[19] The RFs became larger and less topographically organized when transitioning into FST, which is immediately anterior and ventral to MT.[20–22]

Recordings were performed using linear array probes with four or eight tetrodes (NeuroNexus, Inc). Tetrodes were separated by 300 μm. Electrodes within a tetrode were arranged in a diamond pattern and separated by 25 μm. Neuronal signals were sampled at 30 kHz and stored with eye movement and phototransistor traces sampled at 1 kHz (Scout Processor; Ripple, Inc). Tetrode recordings leverage that the amplitudes of recorded action potentials depend on the distance between the neuron and each channel to provide additional waveform information for distinguishing neurons.[68] Tetrode-based spike sorting was performed offline using

the KlustaKwik[69] semi-automatic clustering algorithm in MClust 4.0[70] followed by manual refinement based on waveform, stability, and refractory period violations in Offline Sorter (Plexon, Inc). As in previous work,[34,64] all isolated units were verified by at least two authors. Only neurons for which the stimulus was entirely within the RF were included: 193 MT and 317 FST neurons from the right hemisphere of Monkey J (33 MT and 33 FST sessions); 165 MT and 217 FST neurons from the left hemisphere of Monkey C (12 MT and 20 FST sessions).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Behavioral data analysis

The ability to discriminate toward/away motion was quantified separately for each cue condition by calculating the proportion of 'toward' responses as a function of the directionally signed motion coherence.[8] The data were then fit with a cumulative Gaussian using Psignifit 4 in MATLAB (Mathworks) to estimate the response bias ($\mu$) and sensitivity ($\sigma^{-1}$).[71]

### Visual response latency

The spike times of individual neurons were aligned to the stimulus onset using the phototransistor trace. Following previous work,[34,64] spikes were binned with 1 ms resolution and convolved with a double exponential function, which was averaged over all trials to generate an average spike density function (SDF). The response latency was then defined as the first time point after stimulus onset when the SDF significantly deviated from the average level of baseline activity (computed over all trials using the 150 ms of fixation preceding the stimulus onsets) with maintained significance for at least 30 ms (ANOVA followed by Tukey's HSD test, $p < 0.05$).

### Saccade analysis

A saccade onset was defined as the first time point when the velocity of either eye was $\geq 150^{\circ}$/s. An ANOVA ($p < 0.05$) on mean responses during the 150 ms prior to saccade onsets in the overlap saccade task was used to determine whether a neuron had significant saccade direction selectivity. For all neurons with significant selectivity, trials were grouped according to the preferred and non-preferred saccade direction. For each trial, two SDFs were created aligned to either the saccade target onset or saccade onset. The SDFs were then averaged across neurons to generate population time courses for the preferred and non-preferred saccade directions.

### Asymmetry Index (AI)

Neuronal selectivity for 3D motion was quantified for each cue condition using an Asymmetry Index (AI)[14]:

$$Asymmetry\ Index(AI) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_{T(i)} - R_{A(i)}}{\left|R_{T(i)} - R_{A(i)}\right| + \sigma_{av(i)}} \quad \text{(Equation 1)}$$

where $R_{T(i)}$ and $R_{A(i)}$ are the mean responses to toward and away motion at the i $^{th}$ motion coherence, and $\sigma_{av(i)}$ is the average standard deviation of toward and away responses at that coherence. Responses were computed over the full stimulus duration (1 s). The AI quantifies the degree of asymmetry in the responses to toward versus away motion. Larger magnitudes indicate greater selectivity, and positive (negative) values indicate toward (away) preferences. The statistical significance of toward/away selectivity was determined using a two-way ANOVA for direction and motion coherence ($p < 0.05$).

### Monocularity Index

To interpret visual responses to the combined-cue and stereoscopic cue conditions in relation to monocular views of the same stimuli and ocular dominance, we computed monocularity measures using the monocular trials:

$$Monocularity\ Index = \frac{R_{max} - L_{max}}{R_{max} + L_{max}} \quad \text{(Equation 2)}$$

where $R_{max}$ and $L_{max}$ are the maximum responses during right- and left-eye visual stimulation, respectively. The dominant eye was defined using the sign of the monocularity index, with positive values indicating right dominance and negative values indicating left dominance.

### 2D motion analysis

Selectivity for 2D motion was quantified for each speed and viewing condition (left eye, right eye, or binocular) by performing ANOVAs ($p < 0.05$) on mean firing rates calculated over the full stimulus duration (1 s). Direction preferences were estimated from von Mises fits to the tuning curves.

### Classification of 2D retinal-motion and 3D motion selectivity

To classify neurons as selective for 2D retinal motion or 3D motion, we computed partial correlations for each neuron in 2D retinal-motion and 3D motion coordinates using the left- and right-eye perspective cue responses.

$$R_{2D} = \frac{r_{2D} - r_{3D}r_M}{\sqrt{(1 - r_{3D}^2)(1 - r_M^2)}} \quad \text{(Equation 3)}$$

$$R_{3D} = \frac{r_{3D} - r_{2D}r_M}{\sqrt{(1 - r_{2D}^2)(1 - r_M^2)}} \quad \text{(Equation 4)}$$

Here, $r_{2D}$ is the correlation between the dominant- and non-dominant eyes' perspective cue responses in 2D retinal-motion coordinates and $r_{3D}$ is the correlation between the responses aligned relative to the direction of 3D motion (toward/away). The 2D retinal-motion coordinates were obtained by reflecting the dominant eyes' perspective cue responses about the 0% motion coherence. Lastly, $r_M$ is the correlation between the dominant eye's perspective cue responses in the two coordinate frames. The partial correlation coefficients were then converted to $Z$-scores ($Z_{2D}$ and $Z_{3D}$) using Fisher's $r$ to $Z$ transformation:

$$Z_{2D} = 0.5\ln\left(\frac{1+R_{2D}}{1 - R_{3D}}\right)\sqrt{df} \quad \text{(Equation 5)}$$

$$Z_{3D} = 0.5\ln\left(\frac{1+R_{3D}}{1 - R_{2D}}\right)\sqrt{df} \quad \text{(Equation 6)}$$

Here, the degrees of freedom, $df$, were equal to the number of motion coherences (13) minus 3. For a neuron to be classified as 3D (or 2D) selective, the difference between $Z_{3D}$ and $Z_{2D}$ was required to exceed 1.28 (or vice versa), corresponding to $P = 0.90$, which is a standard criterion used to classify pattern/component selectivity.[23,24,72] Otherwise, the neuron remained unclassified.

### Linear population decoding

We used Fisher's Linear Discriminant to assess how well toward and away motion could be discriminated using the responses of different neuronal subpopulations.[73] For each decoder, we performed repeated random subsample cross-validation, where 20 trials were sampled with replacement from each neuron's response distribution to the combined-cue stimuli (balanced across motion coherences). Decoders were trained and tested on 50% of the sampled trials (balanced across motion coherences) using mean responses over the stimulus duration. This process was repeated 1,000 times and the generalization of each decoder tested using a bootstrap across all cue conditions and motion coherences.

### Neurometric sensitivity

A neurometric function was computed for each neuron and cue condition. Neuronal responses to each cue were grouped according to the direction preference using the sign of the AI. For each motion coherence, an ROC curve was generated using the response distributions to the preferred and anti-preferred motions, and the area under the curve provided the proportion of 'preferred direction reports'.[28,29] The 0% motion coherence trials were compared using identical distributions and were thus always equal to 0.50. After computing proportions for each coherence, a cumulative Gaussian was fit to estimate the neurometric sensitivity.

### Hierarchical model of 3D motion encoding

A model was developed to explore the potential mechanisms underlying the 2D-to-3D visual motion transformation. We simulated responses to 3D motion directions at all combinations of eight azimuths (0° to 315° in 45° steps) and seven elevations (linearly spaced in the sine of the elevation), for a total of 42 unique directions. Each trajectory simulated motion at 80 cm/s (corresponding approximately to the retinal speed in the experiments, ~4.2°/s) centered at a viewing distance of 57 cm. For each trajectory, we ray-traced the corresponding left- and right-eye 2D velocities assuming an interocular difference of 3 cm.

The first layer included 84 units with 2D direction and speed tuning, and either left or right eye ocular dominance (in equal proportions). The tuning function for each unit was described by:

$$R(\theta, s) = DC + G \times OD \times D(\theta) \times V(s) \quad \text{(Equation 7)}$$

where $\theta$ is the 2D retinal-motion direction, $s$ is the retinal speed, $DC$ is an offset (5 Hz), $G$ is the gain (30 Hz), and $OD$ is the ocular dominance (0.6 for the dominant eye and 0.4 for the non-dominant eye). The 2D direction tuning was described by a von Mises function: $D(\theta) = e^{\kappa[\cos(\theta - \mu) - 1]}$, where the mean ($\mu$) and concentration ($\kappa$) define the preference and bandwidth, respectively. The concentration, $\kappa$, was set to 7, corresponding to a tuning bandwidth of ~90°. The speed tuning was described by a log-Gaussian: $V(s) = \exp\left[-\frac{\log\left(\frac{s+s_0}{s_p+s_0}\right)^2}{2\sigma^2}\right]$, where $s_p$ is the preferred speed, $s_0$ is an offset parameter (0.33), and $\sigma$ is the tuning bandwidth (1.16). The values correspond to median values previously reported for MT[74]. For units preferring upward or downward motion, the speed tuning

bandwidth was decreased to 0.5 to produce similar 3D direction tuning bandwidths in the second layer. The velocity preferences of the units were identical for the two eyes. Each unit was maximally responsive to the 2D retinal motion direction and speed corresponding to one of the 42 simulated 3D motion directions, such that each left (right) eye dominant unit had preferences corresponding to the left (right) eye projection of a specific 3D direction.

The second layer consisted of units that integrated the 2D selective responses of the first-layer units. There were two sets of 42 units (left-eye dominant or right-eye dominant; 84 units total) that linearly combined pairs of first-layer units, followed by an expansive nonlinearity: $R(\theta, s) = [OD \times R_L(\theta, s) + (1 - OD) \times R_R(\theta, s)]^{\delta}$, where $\delta$ is the expansive nonlinearity (1.1) and units could be left-eye ($OD = 0.6$) or right-eye ($OD = 0.4$) dominant. By integrating the output of pairs of first-layer units ($R_L$ and $R_R$) whose preferences corresponded to the eye-specific 2D retinal motion projections of the 3D motion, each unit was maximally responsive to one of the 42 3D directions. For example, a second-layer unit that preferred away motion integrated the output of a first-layer, left-eye dominant unit that preferred leftward motion at $\sim$4.2°/s and a first-layer, right-eye dominant unit that preferred rightward motion at $\sim$4.2°/s.

### Vergence control

We used a linear mixed effects model to test if small changes in vergence that did not violate the vergence window might have affected our assessments of 3D selectivity. The analysis was performed on the combined-cue data since these stimuli had the largest number of cues that could elicit changes in vergence. On each trial, we computed the mean vergence over the stimulus duration. For each neuron with significant combined-cue tuning, we then tested if the combined-cue firing rate depended on main effects of motion coherence, motion direction, and vergence. Note that this was the same test used to test for combined-cue selectivity, but with vergence added as a covariate. Only 19/194 (9.8%) MT and 25/309 (8.1%) FST neurons showed a significant effect of vergence ($p < 0.05$). Most importantly, the significance of the main effect of motion direction did not depend on whether vergence was included as a covariate for all but 4 (2.1%) MT and 3 (1.0%) FST neurons. These findings are similar to those in previous reports[8,14,29,34,64,75] and imply that vergence signals did not account for 3D motion selectivity.

**Cell Reports, Volume *42***

## Supplemental information

# Hierarchical computation of 3D motion across macaque areas MT and FST

**Lowell W. Thompson, Byounghoon Kim, Bas Rokers, and Ari Rosenberg**

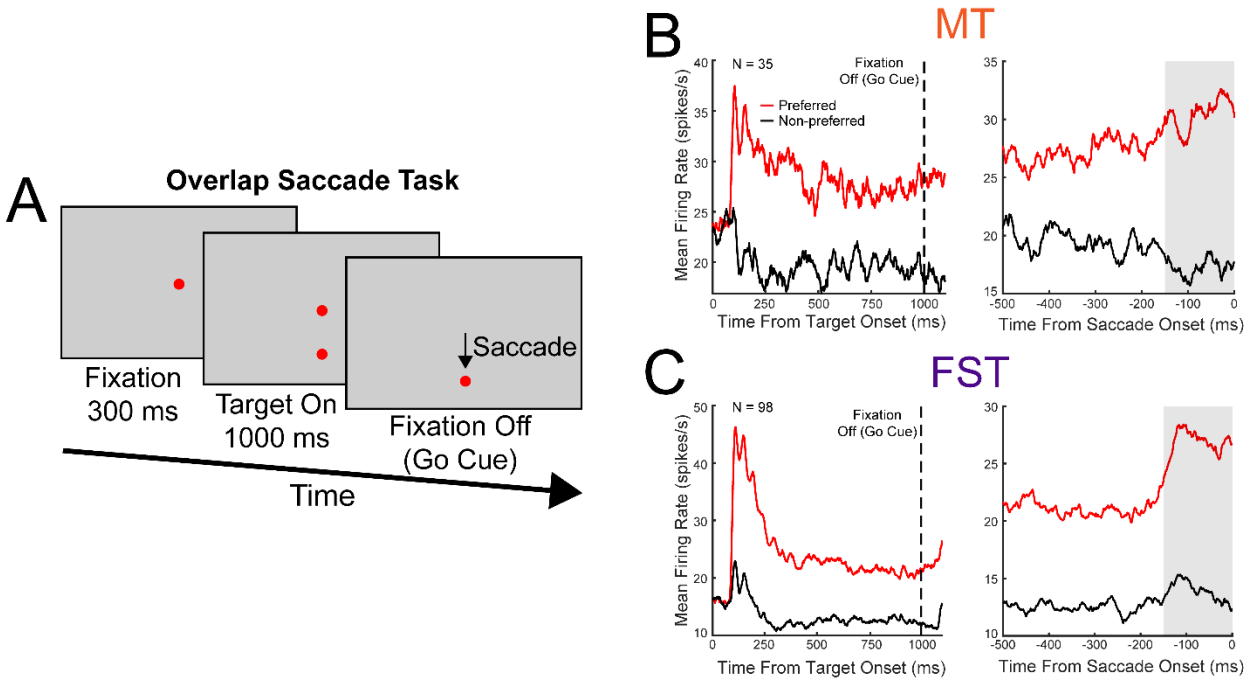



**Figure S1. Saccade-related activity during an overlap saccade task. Related to Figure 2.**
**(A)** Each trial, the monkey fixated a central target for 300 ms before a saccade target appeared either above or below the fixation target for 1 s. The saccade targets had the same locations as the choice targets in the 3D motion discrimination task. When the fixation point disappeared, the animal made a saccadic eye movement to the target.
**(B)** Mean spike density functions (preferred saccade direction: red; non-preferred direction: black) for the 35 MT neurons with significant saccade direction selectivity. The left plot shows the time courses aligned to the saccade target onset. The vertical black dashed line indicates when the fixation point disappeared (go cue). The right plot shows the time courses aligned to the saccade onset. The gray shaded region marks the time segment used to assess saccade selectivity.
**(C)** Same as ***B***, but for the 98 FST neurons with saccade direction selectivity. Compared to MT, FST showed more prominent response modulation beginning ~150 ms before the saccade onset.

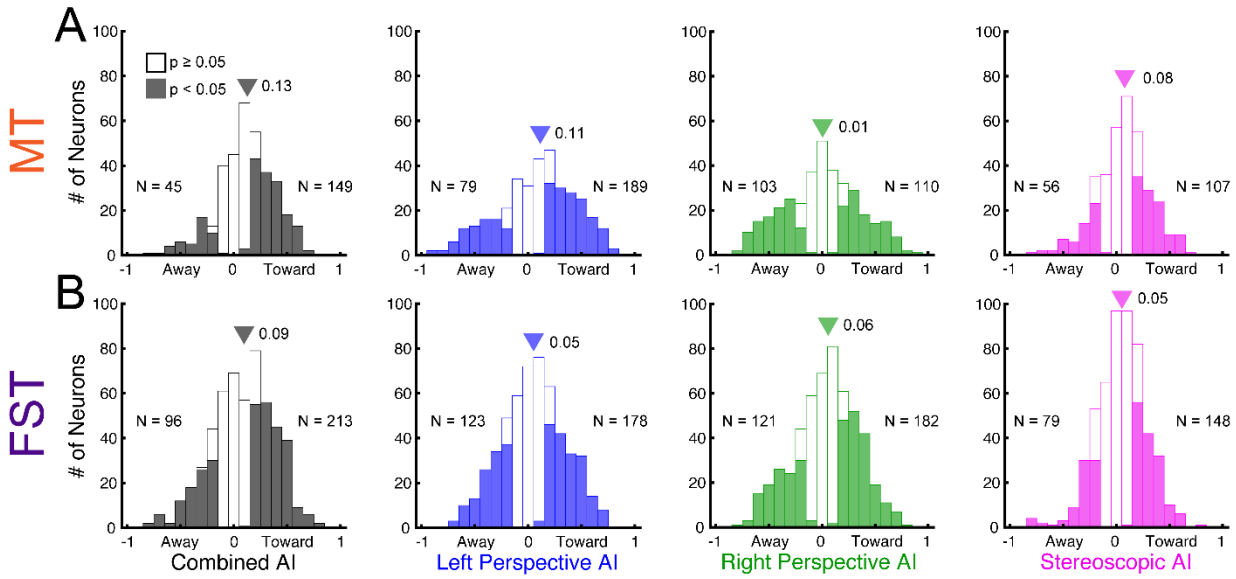



**Figure S2. Distributions of toward/away selectivity for each cue condition. Related to Figure 3.**

**(A)** Distribution of AIs in MT for each cue condition (combined-cue: black; left-eye perspective cues: blue; right-eye perspective cues: green; stereoscopic cues: magenta; N = 358). Filled bars indicate significant toward/away selectivity. Insets indicate the number of neurons with significant toward (positive) or away (negative) selectivity. Arrows mark mean AIs.

**(B)** Same as ***A***, but for FST (N = 534). In both areas, more neurons preferred toward than away motion.

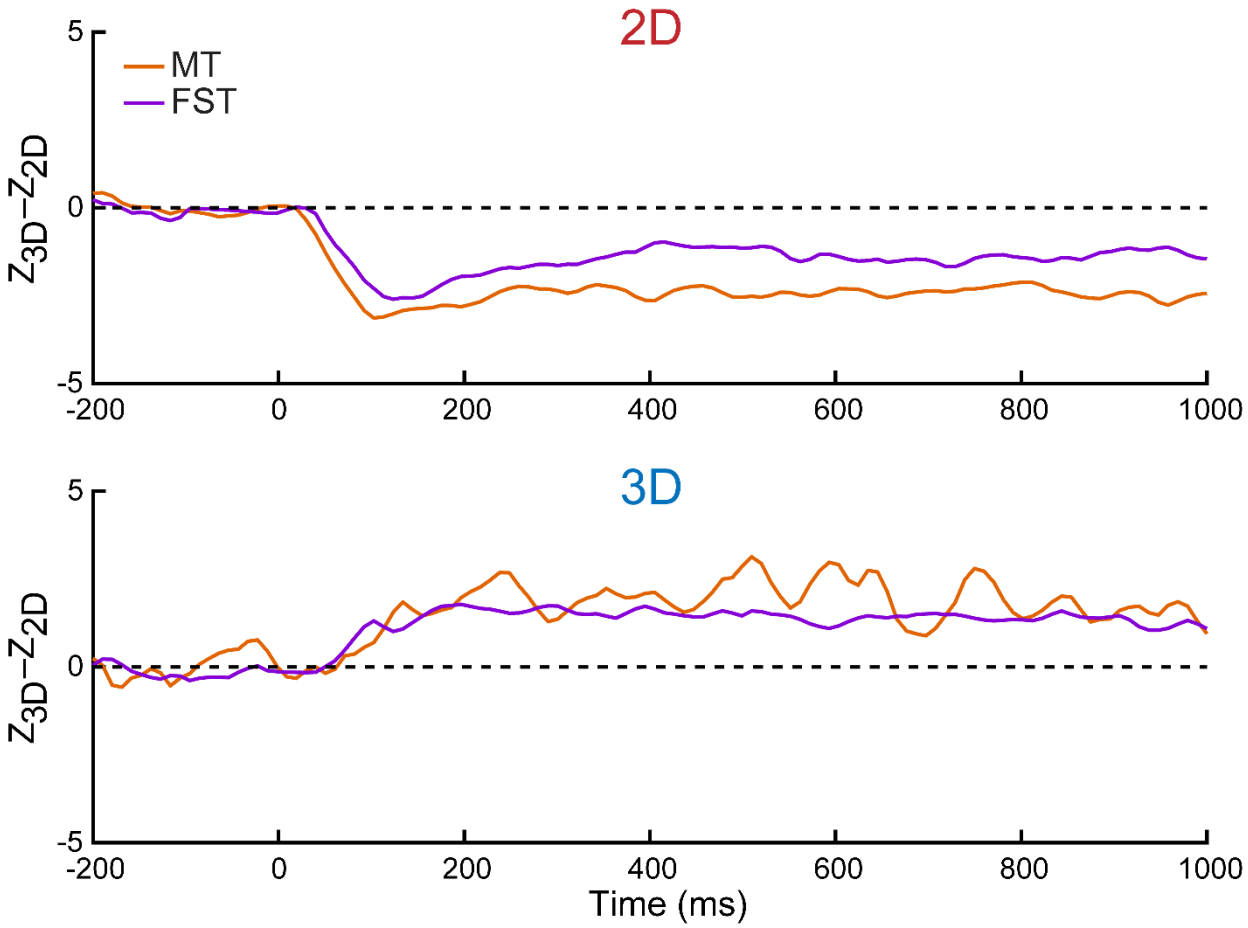



**Figure S3. Time courses of 2D and 3D motion selectivity. Related to Figure 4.**
**(A)** Time courses of $Z_{3D} - Z_{2D}$ averaged over all 2D selective MT (orange) and FST (purple) neurons (50 ms sliding bins with 1 ms increments). The latency difference between the times courses (cross-correlation peak: $\tau$ = 11 ms) was consistent with the areas' visual response latencies (**Figure 2C**). Note that 2D selectivity was stronger in MT than FST throughout the response.
**(B)** Same as ***A***, but for 3D selective neurons. Statistical comparisons were not possible given the small number of 3D selective MT neurons (N = 9), but it appears that 3D selectivity may have first emerged in FST, pointing to a possible role for feedback in MT's representation of 3D motion.

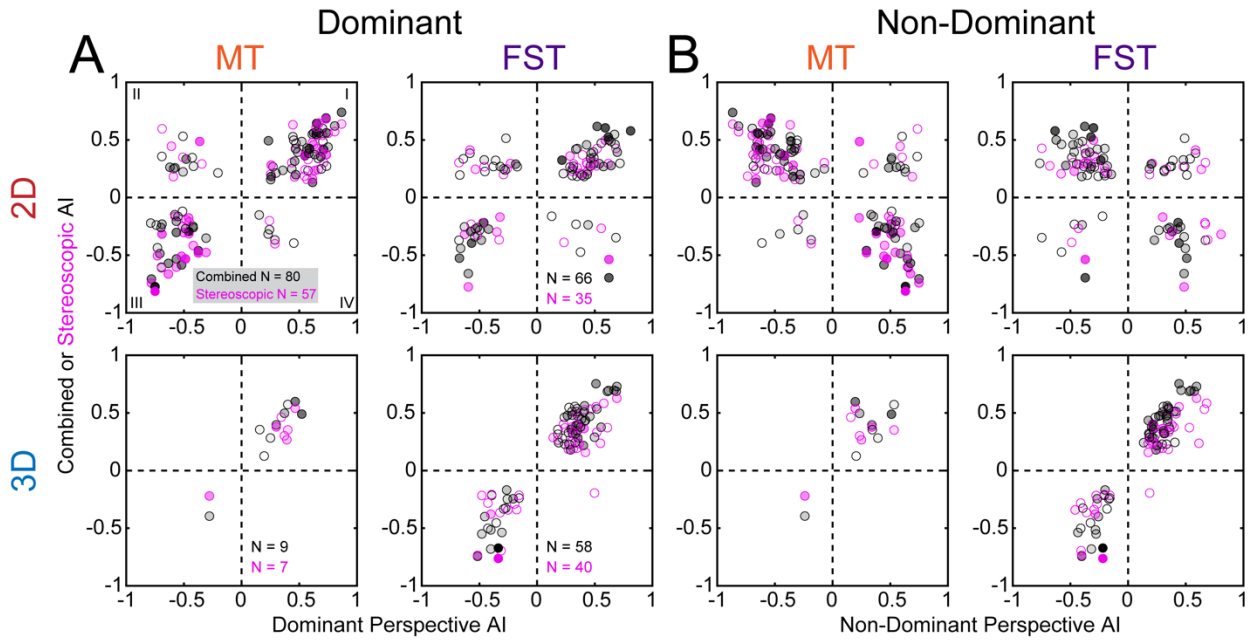



**Figure S4. Comparison of combined-cue and stereoscopic cue AIs with dominant and non-dominant eye perspective cue AIs. Related to Figure 4.**

**(A)** Combined-cue (black points) and stereoscopic cue (magenta points; neurons with significant stereoscopic cue tuning only) AIs versus dominant eye perspective cue AIs for 2D (top) and 3D (bottom) selective MT (left) and FST (right) neurons. Direction preferences were positively associated. More opaque points indicate stronger ocular dominance (normalized for each area), which moderated the strength of the relationships. As expected for 2D selective neurons (top row), data points in the unanticipated quadrants (2 and 4) were generally associated with weak OD.

**(B)** Combined-cue and stereoscopic cue AIs versus non-dominant eye perspective cue AIs. Direction preferences were negatively associated for 2D neurons (top) and positively associated for 3D neurons (bottom). As expected for 2D selective neurons (top row), data points in the unanticipated quadrants (1 and 3) were generally associated with weak OD.

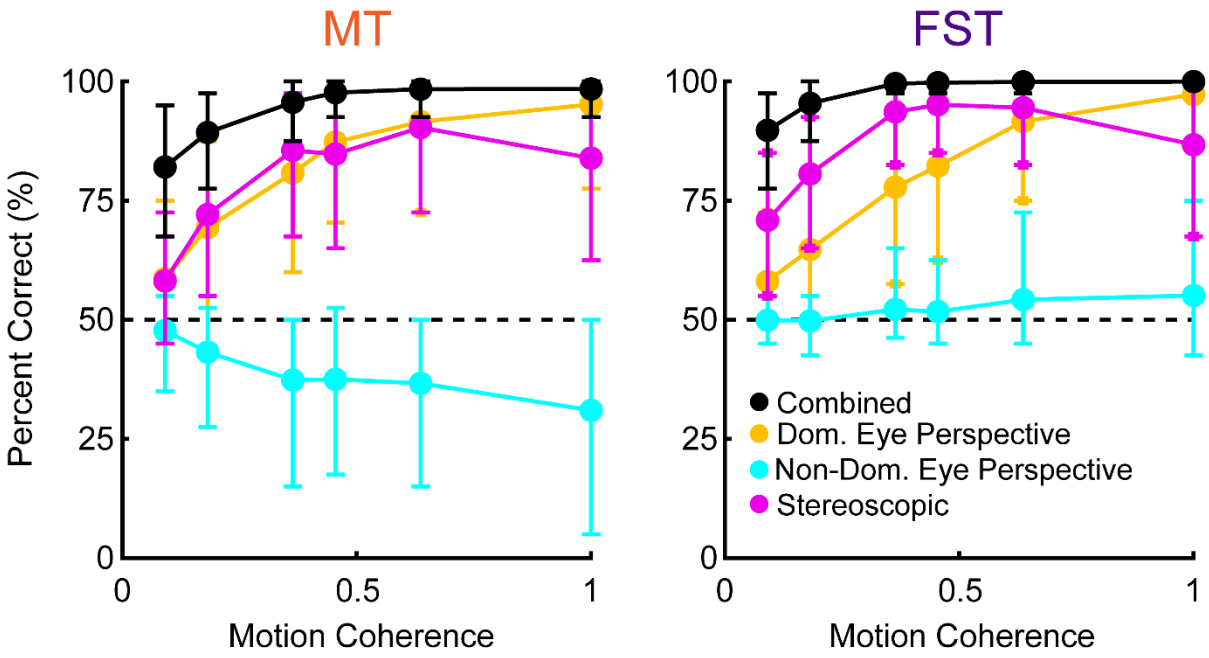



**Figure S5. Linear decoding of 3D motion direction from the full populations. Related to Figure 5.**

**(A)** Decoding performance with all 2D and 3D selective MT neurons as a function of motion coherence. Decoders were trained on the combined-cue responses and tested on the combined-cue (black), stereoscopic cue (magenta), dominant eye perspective cue (yellow), and non-dominant eye perspective cue (cyan) responses. Error bars are 95% confidence intervals.

**(B)** Same as ***A***, but for FST.

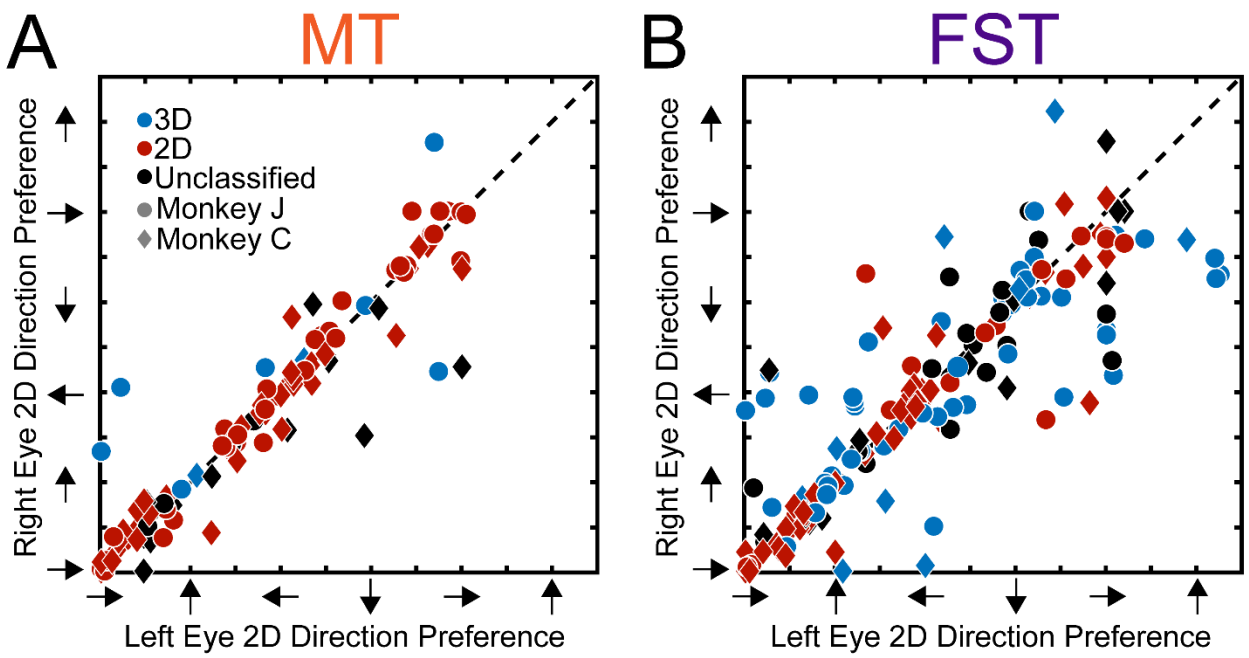



**Figure S6. Interocular 2D direction preference differences. Related to Figure 7.**
**(A & B)** Comparison of left- and right-eye 2D direction preferences in MT (***A***) and FST (***B***). Points are colored according to the classification in **Figure 4**. Dashed black lines are identity lines.

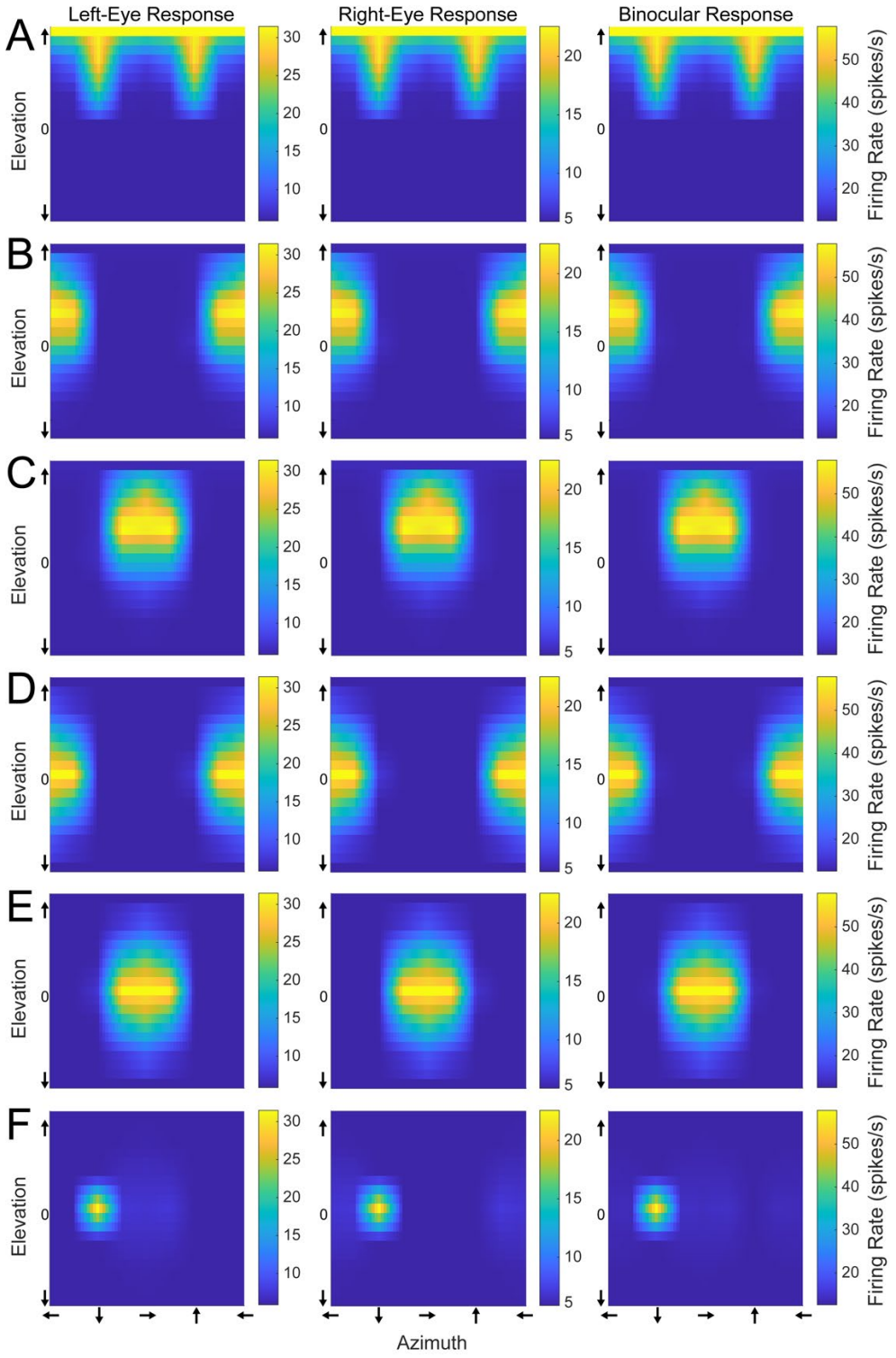



**Figure S7. Responses of example units from the second layer. Related to Figure 7.**
**(A-F)** Responses of six second-layer units to each 3D motion direction, plotted using an equal-area Lambert projection. Yellow hues indicate higher firing rates and blue hues indicate lower firing rates. The units had cue-invariant 3D direction preferences. These examples are a representative subset of the population, with preferences in the northern hemisphere (***A***: upward preference; ***B***: up and to the left; ***C***: up and to the right; ***D***: leftward; ***E***: rightward; ***F***: toward).

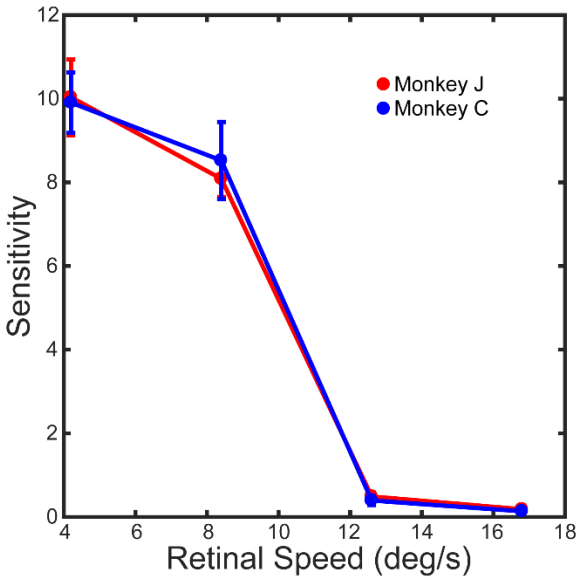



**Figure S8. Behavioral sensitivity to 3D motion as a function of retinal speed. Related to STAR Methods.**

Sensitivity to combined-cue 3D motion decreased at retinal speeds higher than used in the experiments (~4.2°/s).