The role of binocular disparity and attention in the neural representation of multiple moving stimuli in the visual cortex

Anjani Sreeprada Chakrala*, Jianbo Xiao*, Xin Huang

Department of Neuroscience, School of Medicine and Public Health,
University of Wisconsin – Madison

* These authors contributed equally to this work.

Correspondence should be addressed to:

Xin Huang, Department of Neuroscience, University of Wisconsin, Madison, WI 53705, USA
Email: Xin.Huang@wisc.edu

Keywords: neural coding, encoding, MT cortex, binocular disparity, transparent motion, divisive normalization, segmentation, surface-based and object-based attention, macaque monkey.

Acknowledgment

We thank Dr. Greg DeAngelis for sharing the data set of the disparity tuning curves of a large number of MT neurons from his lab to be used in our model, Bryce Arseneau for animal training, and Drs. Jennifer Coonen and Kevin Brunner at the Wisconsin National Primate Research Center for excellent veterinary care and surgical assistance.

Grant support

Research reported in this publication was supported by the National Eye Institute of the National Institutes of Health under Award Number R01EY022443. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest

None.

Abstract

Segmenting visual scenes into distinct objects and surfaces is a fundamental visual process, with stereoscopic depth and motion serving as crucial cues. However, how the visual system uses these cues to segment multiple objects is not fully understood. We investigated how neurons in the middle-temporal (MT) cortex of macaque monkeys represent overlapping surfaces at different depths, moving in different directions. Neuronal activity was recorded from three male monkeys during discrimination tasks under varying attention conditions. We found that neuronal responses to overlapping surfaces showed a robust bias toward the binocular disparity of one surface over the other. The disparity bias of a neuron was positively correlated with the neuron's disparity preference for a single surface. In two animals, neurons preferring near disparities of single surfaces (near neurons) showed a near bias for overlapping stimuli, while neurons preferring far disparities (far neurons) showed a far bias. In the third animal, both near and far neurons displayed a near bias, though the near neurons showed a stronger near bias. All three animals exhibited an initial near bias across neurons relative to the average of the responses to the individual surfaces. Although attention modulated neuronal responses, the disparity bias was not caused by attention. We also found that the effect of attention was consistent with object-based, rather than feature-based attention. We proposed a model in which the pool size of the neuron population that weighs the responses to individual stimulus components can be variable. This model is a novel extension of the standard normalization model and provides a unified explanation for the disparity bias across animals. Our results reveal how MT neurons encode multiple stimuli moving at different depths and present new evidence of response modulation by object-based attention. The disparity bias allows subgroups of neurons to preferentially represent individual surfaces of multiple stimuli at different depths, thereby facilitating segmentation.

Introduction

Natural scenes often contain multiple entities within a three-dimensional (3D) space. Segmenting these visual scenes into distinct objects and surfaces is a fundamental process of vision. Among various visual features, stereoscopic depth and motion cues are particularly important for distinguishing different objects from one another and segregating figural objects from their background (Gibson, 1950; Wagemans et al., 2012). While the visual system can rely on depth or motion cues alone for segmentation, combining these cues further enhances the process (Snowden and Rossiter, 1999; Goutcher 2016). For example, it is easier to segment overlapping surfaces moving in different directions when they are at different depths rather than at the same depth (Qian et al., 1994; Hibbard and Bradshaw, 1999; Greenwood and Edwards, 2006). To understand the neural basis of natural vision, it is important to determine how the visual system represents multiple visual stimuli that differ in various features and how depth cues facilitate the segmentation of multiple stimuli.

The middle-temporal cortex (MT) of primates is crucial for processing motion and depth information (Born and Bradley, 2005; Britten, 2003; Pasternak and Tadin, 2020). Most MT neurons are selective to motion direction and binocular disparity (Maunsell and van Essen, 1983a, 1983b; Albright, 1984; DeAngelis et al., 1998; DeAngelis and Newsome, 1999). Although the representation of binocular disparity for a single surface by MT neurons is well understood (Cumming and Parker, 1999; DeAngelis and Uka, 2003; Uka and DeAngelis, 2003, 2004; Cumming and DeAngelis, 2001), the understanding of how MT neurons represent multiple stimuli at different disparities remains incomplete.

One previous study showed that the responses of MT neurons to two overlapping stimuli moving in opposite (the preferred and null) directions became stronger as the disparity difference between the stimuli increased, suggesting reduced opponent inhibition between motion signals with increased disparity difference (Bradley et al., 1995). Other studies showed that MT neurons can represent structure-from-motion (Bradley et al., 1998; Dodd et al., 2001; Krug et al., 2013) and relative disparity (Krug and Parker, 2011) using overlapping surfaces moving in opposite directions at different disparities. Uka and DeAngelis (2006) also investigated MT neurons' selectivity for relative disparity using center-surround stimuli moving in the same (preferred) directions of the neurons at different disparities. However, to gain a general understanding of how MT neurons encode multiple stimuli at different disparities, it is essential to examine visual stimuli across different motion directions and with varying angular separations between them. We address this question by measuring the direction tuning curves of MT neurons to two overlapping surfaces located at different disparities – one near and one far from the fixation plane – moving in various directions across 360° with different angular separations.

We anticipate that MT neurons will follow one of three encoding rules for multiple stimuli at different disparities. First, they may average the responses to individual stimulus components at different disparities, treating both components equally. Second, they may show a consistent bias toward one stimulus disparity (e.g. the near surface), regardless of individual neurons' disparity preferences for single surfaces, which could be beneficial for efficiently representing disparity information or performing tasks like figure-ground segregation in natural environments (Sprague et al., 2015; Manning et al., 2024). Third, MT neurons may bias their response toward the disparity

component closer to their preferred disparity, enhancing the extraction of information from the more preferred disparity component. The primary goal of this study is to determine the encoding rules for multiple disparities of moving stimuli.

Attention can bias neuronal responses elicited by multiple two-dimensional (2D) stimuli in the receptive field (RF), in favor of the attended stimulus (Desimone and Duncan, 1995; Reynolds et al., 1999; Treue and Maunsell, 1996; Ferrera and Lisberger, 1997; Recanzone and Wurtz, 2000). MT neuron responses can be influenced by feature-based attention, such as attention to motion direction (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004) and binocular disparity (Ruff and Born, 2015). Attention can also select a specific visual object (Duncan, 1984; Cavanagh et al., 2023; Valdes-Sosa et al., 1998, 2000; Mitchell et al., 2003, 2004). Wannig et al. (2007) demonstrated that MT neuron responses could be modulated by object-based attention directed at a moving surface composed of two overlapping 2D stimuli. However, the effects of selective attention on the neural representation of multiple stimuli located at different depths remain underexplored. When subjects are instructed to attend to the disparity of one of two overlapping moving surfaces at different depths, it is unclear whether the effect of attention is feature-based and restricted to disparity or if it extends to all the features of the entire surface, indicating object-based attention. This study aims to address this question as its secondary goal.

We recorded neuronal activities in MT of three macaque monkeys while they performed discrimination tasks under different attention conditions. We found that neuronal responses to overlapping surfaces at different depths showed a robust bias toward the disparity of one of the surfaces. Across all animals, a neuron's disparity bias in response to two surfaces was positively correlated with its disparity preference for a single surface. Notably, the combined responses of near and far neurons of all three animals showed an initial bias toward the near surface. Additionally, we found that the disparity bias was not caused by attention, although attention modulated neuronal responses in a manner consistent with object-based attention. We proposed a model that extended the standard normalization model to account for our findings.

Results

We asked how the primate visual system utilizes disparity and motion cues to segment multiple visual stimuli. To address this question, we characterized how neurons in area MT of macaque monkeys represent overlapping stimuli located at different depths and moving simultaneously in different directions, while monkeys performed discrimination tasks under different attention conditions.

Behavioral task and performance

We trained two monkeys (B and G) to perform a direction discrimination task. The visual stimuli were two overlapping random-dot patches moving in two directions separated by either 60° or 120°. By randomly interleaving trials with 60° and 120° direction separations, monkeys are prevented from performing the task by merely identifying the direction of one component and guessing the other based on a fixed separation. The two direction separations, 120° and 60°, also offered relatively easier and more challenging scenarios for segmentation, respectively. One patch was located at a near disparity (-0.1°) and the other patch at a far disparity (0.1°). The monkey's task was to report the motion direction of the patch at a cued depth (either the near or far disparity) by making a saccadic eye movement to one of twelve reporting targets (12 AFC task) (Fig. 1A, B). We varied the vector average (VA) direction of the overlapping stimuli across trials. The cue-near trials (Fig. 1A) and cue-far trials (Fig. 1B) were randomly interleaved. To perform the task well, the monkeys need to segment two overlapping surfaces and attend to the cued surface to discriminate its motion direction. We therefore refer to the cue-near and cue-far trials as "Attend Near" and "Attend Far" trials, respectively. We also randomly interleaved trials in which only one of the two patches was presented at either the near (-0.1°) or far (0.1°) disparity.

We recorded neuronal activities in area MT while the monkeys performed the task (49 experimental sessions for monkey B, 44 sessions for monkey G). Both monkeys performed the task well. The mean correct rate across sessions was 75% (sd = 6.6) for monkey B (single-surface 87%, two-surface 72%), and 72% (sd = 8.1) for monkey G (single-surface 87%, two-surface 68%). As expected, both monkeys performed significantly better in the single-surface trials than in two-surface trials (signed rank test, $p < 10^{-8}$) (Fig. 1C, E). Notably, while monkey B performed equally well in the "Attend Near" and "Attend Far" trials (signed rank test, $p = 0.56$) (Fig. 1C, D), monkey G performed significantly better at discriminating the direction of the near-surface (mean = 72% correct) than the far-surface (mean = 65% correct) of the overlapping stimuli (one-sided signed rank test, $p= 4.3\times10^{-13}$) (Fig. 1E, F).

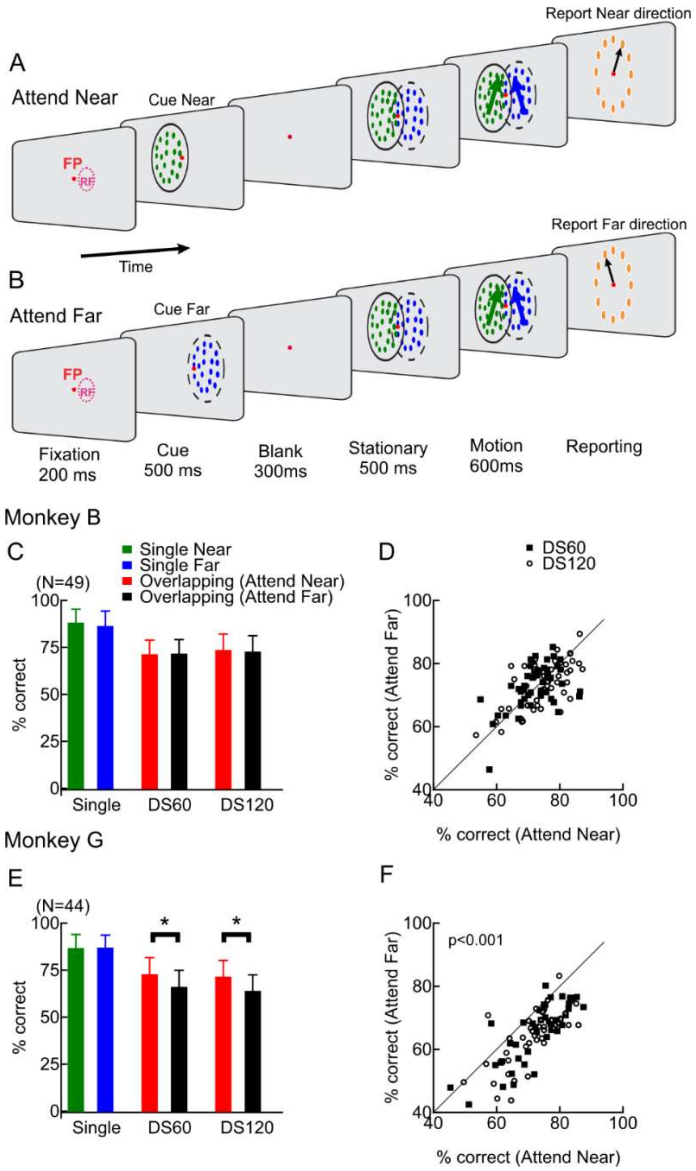


Figure 1. The behavioral task and performance. A, B. Visual stimuli were overlapping random-dot patches located at different depths and moving in different directions. Each trial started with fixation on a spot at the center of the monitor. Then, as a cue, a stationary random-dot patch (RDP) was presented for 500 ms at either the near (-0.1°) (A) or the far (0.1°) disparity (B) for 500 ms. After a 300-ms blank period, overlapping random-dot stimuli were turned on, initially stationary for 500 ms, and then moved in two directions for 600 ms. The direction separation (DS) between the two motion directions was either 60° or 120°. After the visual stimuli were turned off, 12 reporting targets were illuminated. To receive a juice reward, the monkey needed to make a saccadic eye movement to one of the targets in the direction that matched the motion direction of the cued surface of overlapping stimuli. **A.** Cue-near (Attend-Near) trial. **B.** Cue-far (Attend-Far) trial. **C. D.** Behavioral performance of monkey B. C shows the mean correct percentages across sessions for single and overlapping stimuli. D compares the performance between Attend Near and Attend Far conditions. Each dot in D represents the trial-averaged correct percentage in one session. **E. F.** Behavioral performance of monkey G, following the same convention as in C and D. Error bars in C and E represent standard deviation. (*) in E indicates $p < 0.001$ (paired signed-rank test).

Direction tuning of example MT neurons to overlapping stimuli moving at different depths

Figure 2 shows the direction tuning curves of four example MT neurons to overlapping stimuli moving in different directions at different depths. The red and black curves indicate the responses under attend-near and attend-far conditions, respectively. The overlapping stimuli had four configurations (Fig. 2A-D). The motion direction at the near surface could either be at the clockwise (A, C) or counter-clockwise side (B, D) of the two motion directions, and the direction separation between the two directions was either 60° (A, B) or 120° (C, D). We referred to the overlapping stimuli as the bi-directional stimuli. Figure 2 also shows the direction tuning curves to single motion directions of the constituent near-surface (green) and far-surface (blue) when they were presented individually. We classified a neuron as a near-preferred (or far-preferred) neuron if the maximum response of the neuron's direction tuning curve to the near (or far) component was greater than that to the other component.

Figure 2A1-D1 shows the responses of a near-preferred neuron, referred to as "near neuron", from monkey B. The response tuning curve to the bi-directional stimuli under the attend-near condition (red curve) was biased toward the near component (green curve). By definition, the near neuron responded more strongly to the near component than the far component, and therefore the average of the component responses (gray) had a peak toward the near component. However, the red curve showed more bias toward the near component than expected by the component average. This near bias was found at both direction separations of 60° and 120°, regardless of whether the near direction was at the clockwise or counterclockwise side of the two motion directions. Under the attend-far condition, the response tuning to the bi-directional stimuli still showed a near bias in most of the stimulus configurations (black curve, Fig. 2A1, C1, D1), but to a lesser degree than that under the attend-near condition.

Figure 2A2-D2 shows the responses of a far neuron from monkey B. The response tuning curve to the bi-directional stimuli under the attend-far condition (black curve) was biased toward the far component (blue curve). Under the attend-near condition, the response tuning to the bi-directional stimuli still showed a far bias in some stimulus configurations, albeit weaker (red curve, Fig. 2B2, C2, D2).

Figures 2A3-D3 and 2A4-D4 show the responses of an example near neuron and a far neuron from monkey G. The near neuron showed a strong near bias under either attend-near or attend-far conditions (Fig. 2A3-D3). Remarkably, the far neuron also showed a near bias regardless of the attention (Fig. 2A4-D4). The responses to the bi-directional stimuli of these two neurons showed small differences between the attend-near and attend-far conditions.

We previously found that some MT neurons showed response bias toward one direction component of two overlapping stimuli moving in different directions within a 2-dimensional plane (Xiao and Huang, 2015). We noticed that some MT neurons also showed this directional bias to stimuli moving at different depths. For example, the far neuron from monkey B showed a stronger far bias when the far direction was at the counter-clockwise side of the two motion directions (the black curve was very close to the blue curve, Fig. 2A2, C2) than when the far direction was at the clockwise side (Fig. 2B2, D2). Our experimental design allowed us to separate the effect of disparity bias from the directional bias as shown below.

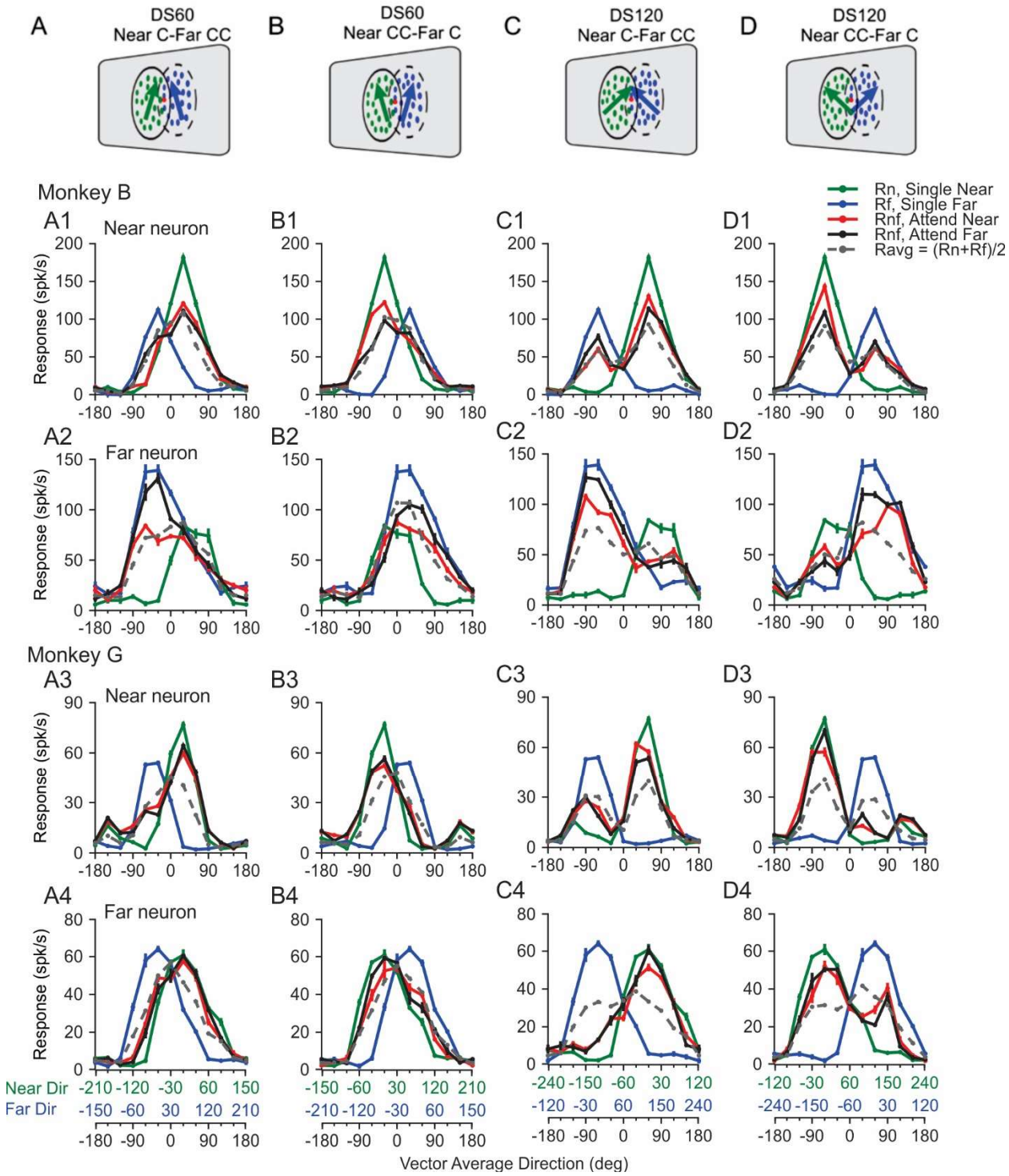

A
DS60
Near C-Far CC
B
DS60
Near CC-Far C
C
DS120
Near C-Far CC
D
DS120
Near CC-Far C
Monkey B
Rn, Single Near
Rf, Single Far
Rnf, Attend Near
Rnf, Attend Far
Ravg = (Rn+Rf)/2
A1
Near neuron
B1
C1
D1
A2
Far neuron
B2
C2
D2
Monkey G
A3
Near neuron
B3
C3
D3
A4
Far neuron
B4
C4
D4
Response (spk/s)
Near Dir
Far Dir
Vector Average Direction (deg)

Figure 2. Direction tuning curves of example MT neurons in response to overlapping bi-directional stimuli and the constituent stimulus components. A-D. Stimlus configurations. The Direction separation between the two stimulus components was either 60° (A, B) or 120° (C, D). The motion direction at the near surface can either be at the clockwise (A, C) or counter-clockwise side (B, D) of the two motion directions. Responses of a near neuron **(A1-D1)** and a far neuron **(A2-D2)** from monkey B. Responses of a near neuron **(A3-D3)** and a far neuron **(A4-D4)** from monkey G. The red and black curves indicate the responses under attend-near and attend-far conditions, respectively. The green and blue curves indicate the responses to single motion directions of the near-surface and far-surface, respectively, when they were presented individually. The gray dash line indicates the average of the responses to the component directions (i.e. average of the green and blue curves). The abscissas in green and blue show the directions for the near component and the far component, respectively, of the corresponding bi-directional stimuli for which the vector average (VA) direction is shown by the black abscissa. The green and blue axes are shifted by 60° relative to each other in columns A and B, and by 120° in columns C and D. VA direction of 0° is aligned with each neuron's preferred direction (PD). Error bars indicate the standard error.

Population-averaged direction tuning curves

Figure 3 shows the tuning curves to the bi-directional stimuli averaged across near neurons and far neurons for each animal. The direction tuning curve of each neuron was first fitted with a spline, normalized, and then shifted circularly such that the neuron's preferred direction (PD) was aligned with the VA direction 0° of the bi-directional stimuli. To examine disparity bias and balance potential directional bias, we averaged the response tuning curves to the bi-directional stimuli when the near component was at the clockwise side (Near C/Far CC) (e.g. Fig. 2A1) and at the counter-clockwise (Near CC/Far C) side (e.g. Fig. 2B1) of the two component directions. Before averaging, we flipped the tuning curve of the stimulus configuration when the near component was at the counter-clockwise side (Near CC/Far C) relative to VA 0°, so in the flipped tuning curve, the near component was also at the clockwise side of two component directions. Averaging these two tuning curves preserved the disparity bias but cancelled out the directional bias. As a result, a response bias toward positive VA values indicates a bias toward the near component. The tuning curves of each neuron were then averaged across neurons for 60° (Fig. 3A, B) and 120° (Fig. 3C, D) direction separations.

For monkey B, the population-averaged tuning curves to the bi-directional stimuli showed a bias toward neurons' preferred disparity – the near neurons showed a bias toward the near component (Fig. 3A1, C1), and the far neurons showed a bias toward the far component (Fig. 3B1, D1). The disparity bias was modulated by attention – the near neurons showed a stronger near bias when attention was directed to the near surface (Fig. 3A1, C1) and the far neurons showed a stronger far bias when attention was directed to the far surface (Fig. 3B1, D1). However, the disparity bias cannot be explained by attention modulation. For example, far neurons still showed a clear far bias even when attention was directed to the near surface (red curve in Fig. 3D1).

In contrast, for monkey G, the population-averaged tuning curves to the bi-directional stimuli showed a robust bias toward the near component for both the near neurons (Fig. 3A2, C2) and the far neurons (Fig. 3B2, D2). Even in trials when attention was directed to the far surface, the far neurons still showed a bias toward the near component (black curves in Fig. 3B2, D2). In

comparison to the attention modulation found in the data from monkey B, the effect of attention modulation found from monkey G was small.

Given the apparent discrepancy in the response tuning curves and the magnitude of attention modulation between monkeys B and G, we recorded from area MT of a third monkey R. We asked whether MT neurons showed disparity biases without the involvement of attention. We trained monkey R to perform a demanding fine direction-discrimination task in the visual hemifield opposite from the RFs of the recorded MT neurons, while we presented the bi-directional stimuli in the RFs. The results from monkey R showed a similar disparity bias as found in monkey B – the near neurons tended to show a near bias (Fig. 3A3, C3) and the far neurons tended to show a far bias (Fig. 3B3, D3).

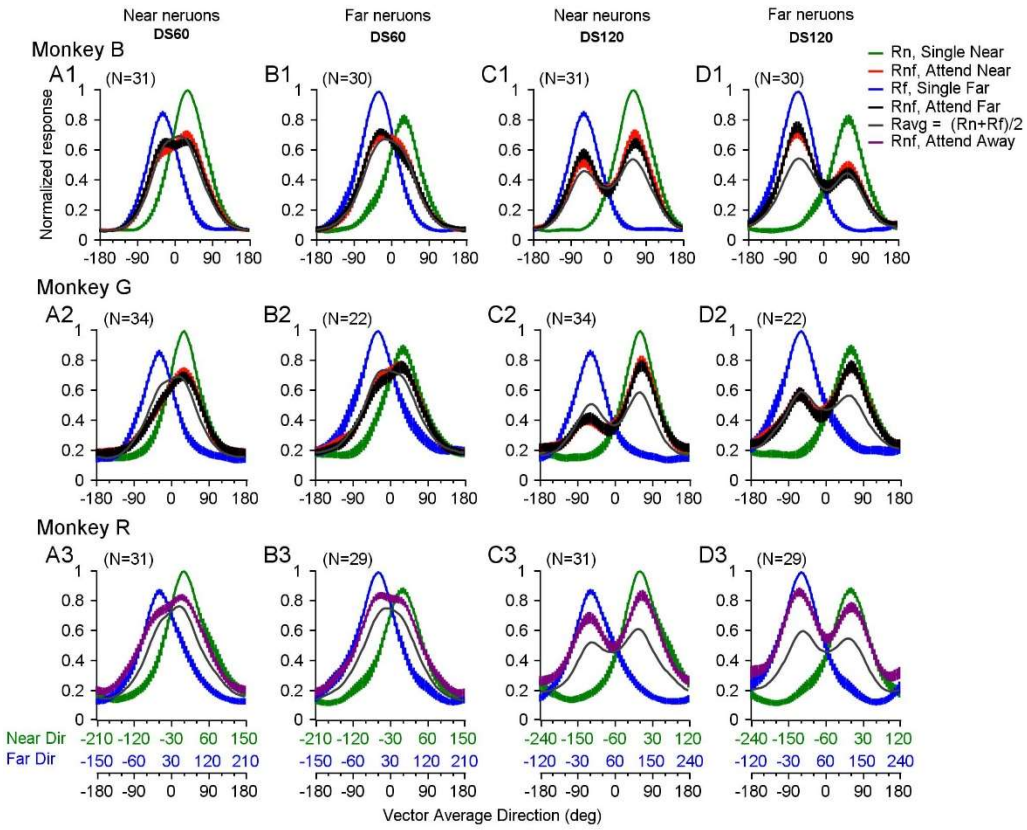


Figure 3. Population-averaged direction tuning curves to overlapping stimuli moving at different depths. Tuning curves to the bi-directional stimuli with 60° direction separation (DS60) (**A, B**) and 120° direction separation (DS120) (**C, D**), and the corresponding component directions. **A, C.** Tuning curves averaged across near neurons. **B, D.** Tuning curves averaged across far neurons. **A1-D1.** Responses from monkey B. **A2-D2.** Responses from monkey G. **A3-D3.** Responses from monkey R when attention was directed away from the RFs. The width of the curve for *Rn*, *Rf*, and *Rnf* represents the standard error. The tuning curve of each neuron was spline-fitted, normalized, and shifted circularly such that each neuron's PD was aligned with VA 0°. The tuning curves to the bi-directional stimuli that had the near component at the clockwise side (Near C/Far CC) and at the counter-clockwise side (Near CC/Far C) of the two component directions were averaged. The tuning curves were then averaged across neurons. As a result, in all panels, the near component is always at the clockwise side of the two motion directions.

Quantifying the disparity bias and the effect of attention

To quantify the response bias toward either the near or far component, we fitted the tuning curve to the bi-directional stimuli of each neuron (the red and black curves illustrated in Fig. 2) as a function of the neuron's responses to the individual stimulus components (the green and blue curves illustrated in Fig. 2). We found that a model of weighted summation plus a nonlinear interaction term, referred to as the SNL model (Eq. 1), provided a good fit of the MT responses. The percentage of variances (PV) (Eq. 13 in Methods) accounted for by the SNL model were 93%, 83%, and 88% for monkeys B, G, and R, respectively. The SNL model fit was significantly better than a linear weighted summation model replacing the nonlinear interaction term with a constant (one-sided signed-rank test, $p < 10^{-7}$).

$$R_{pred}(\theta_n, \theta_f) = w_n R_n(\theta_n) + w_f R_f(\theta_f) + b R_n(\theta_n) R_f(\theta_f) \quad , \qquad (1)$$

in which R_{pred} is the model-predicted response to bi-directional stimuli; θ_n and θ_f are the motion directions of the near and far components; R_n and R_f are the responses to single motion directions at the near and far surfaces; w_n and w_f are the response weights for the near and far components, respectively; b is a coefficient for the multiplicative interaction between the two component responses; w_n, w_f, and b are model parameters.

Using the fitted response weights for the near and far components, we defined a weight bias index (*WBI*) as:

$$WBI = (w_n - w_f)/(w_n + w_f). \qquad (2)$$

A positive *WBI* indicates a bias toward the near component, whereas a negative *WBI* indicates a bias toward the far component. Figure 4 shows the *WBI* of the near and far neurons. In the scatter plots (Fig. 4A, B), each neuron contributed to four points, from direction separations of 60° and 120°, and stimulus configurations of Near C/Far CC and Near CC/Far C. For monkey B, near neurons had significantly larger *WBI* than the far neurons, regardless of whether attention was directed to the near or far surface (one-sided signed-rank test, $p < 0.001$) (Fig. 4A). When attention was directed to the near surface, near neurons showed a significant near bias (mean *WBI* = 0.12, $p = 1.43\text{x}10^{-8}$, one-sided signed-rank test), and far neurons showed a significant far bias (mean *WBI* = -0.17, $p = 1.95\times10^{-8}$). When attention was directed to the far surface, far neurons showed a significant far bias (mean *WBI* = -0.30, $p= 5.74\times10^{-15}$), and near neurons had a mean *WBI* of -0.01 which was marginally less than zero ($p = 0.0537$). Attention had a significant effect on *WBI,* giving rise to a larger *WBI* when attention was directed to the near surface than the far surface across near neurons ($p = 5.4\times10^{-9}$), far neurons ($p = 0.007$), and all neurons combined ($p=3\times10^{-10}$) (Fig. 4A).

For monkey G, both near and far neurons showed a significant near bias, regardless of whether the near surface or the far surface was cued for attention (*WBI* significantly greater than 0, $p = 6.09\text{x}10^{-56}$). Notably, near neurons showed a significantly larger *WBI* than the far neurons when attention was directed to the near surface (one-sided rank sum test, $p < 0.01$) and when the attending near and attending far trials were pooled (one-sided rank sum test, $p = 0.002$). For

monkey G, attention did not change *WBI* significantly across the near neurons (one-sided signed-rank test, p = 0.29), the far neurons (p = 0.59), and all neurons combined (p = 0.41).

For monkey R, we found a similar result that the near neurons showed a significantly larger *WBI* than the far neurons, when attention was directed away from the RFs (one-sided rank sum test, p= 3.94×10^{-5}) (Fig. 4C). Near neurons showed a significant near bias (mean *WBI* = 0.07, p = 4.3×10^{-4}) and far neurons showed a significant far bias (mean *WBI* = -0.03, p = 0.003). In summary, MT neurons in monkeys B and R showed a significant bias toward the stimulus component moving at the more preferred disparity of the neurons. Although neurons in monkey G showed a predominant near bias, near neurons tended to show a stronger near bias than the far neurons.

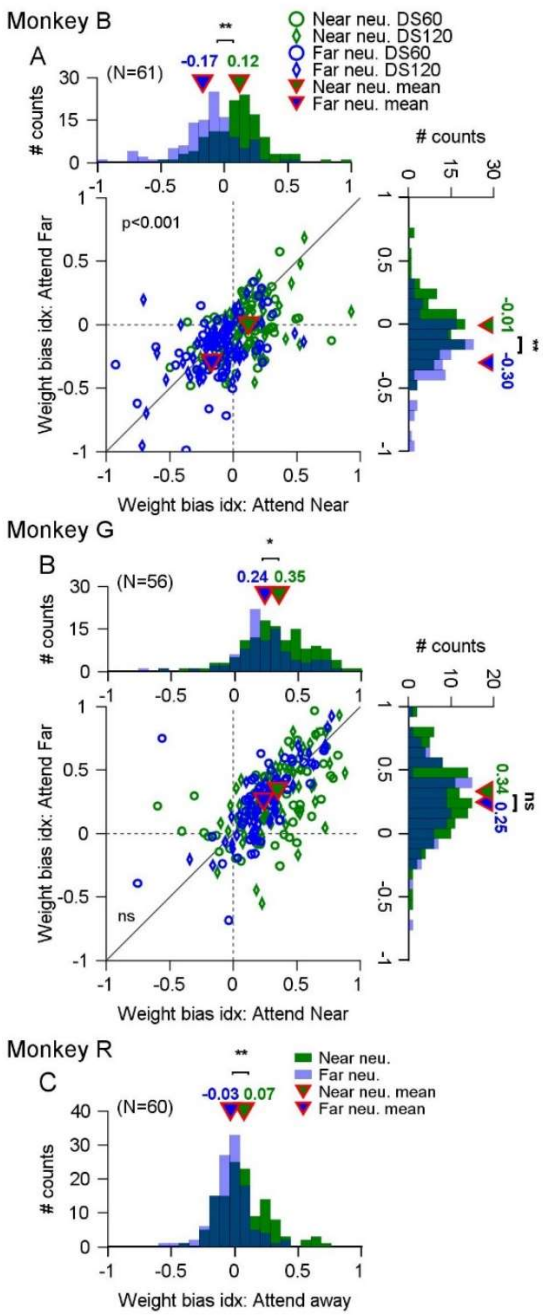


Figure 4. Disparity bias of near and far neurons and the effect of attention. A. B. Scatter plots and marginal distributions of weight bias indices (*WBI*) of near (green) and far (blue) neurons under attend-near and attend-far conditions, for monkeys B (**A**) and G (**B**). In the scatter plots, each point represents the *WBI* based on one bi-directional tuning curve from one neuron. Each neuron contributed to four points, from direction separations of 60° (circle) and 120° (diamond), and stimulus configurations of Near C/Far CC and Near CC/Far C. The *p*-value in the scatter plot was from a paired signed-rank test. The histograms show the marginal distributions of *WBI* in the attend-near (top) and attend-far (right) conditions, for near (green) and far (blue) neurons. **C.** The distributions of *WBI* for monkey R when attention was directed away from the RFs, for near (green) and far (blue) neurons. In all histograms, (**) indicates p < 0.001, (*) indicates p < 0.01, (ns) indicates p > 0.05 (one-sided signed-rank test). The triangles show the mean *WBI* values for near and far neurons. The p value (A) and ns (B) in the scatter plots indicate significance level of comparing *WBI* between attending near and attending far conditions for all neurons combined.

Relationship between disparity bias in response to two surfaces and disparity preference to a single surface

We asked whether a neuron's disparity bias toward one depth of two overlapping surfaces was related to the neuron's disparity preference to a single surface. To quantify the disparity bias toward one depth of two surfaces, we defined a two-surface disparity bias index (*TDBI*) (Eq. 3).

$$TDBI = \frac{(w_n - w_f)_{nC} + (w_n - w_f)_{nCC}}{(w_n + w_f)_{nC} + (w_n + w_f)_{nCC}}, \quad (3)$$

TDBI was similar to the weight bias index (*WBI*) (Eq. 2), except that *TBI* considered both stimulus configurations of Near C/Far CC and Near CC/Far C, and pooled the response weights obtained when the near component moved in the clockwise (*nC*) and counterclockwise (*nCC*) side of two motion directions. By doing so, *TDBI* balanced the weight difference between the near and far components contributed by potential directional bias, and therefore reflected the disparity bias. A positive *TDBI* indicates a disparity bias toward the near component of two surfaces and a negative *TDBI* indicates a bias toward the far component.

We also quantified a neuron's preference for the disparity of a single surface, measured by a single-surface disparity preference index (*SDPI*):

$$SDPI = \frac{\max(R_n) - \max(R_f)}{\max(R_n) + \max(R_f)}, \quad (4)$$

in which R_n and R_f were the direction tuning curves to the single near and far surfaces, respectively, when they were presented alone. The data samples from monkeys B and R were roughly balanced between near and far neurons. The neurons from monkeys B and R had median *SDPIs* of 0.0095 and 0.0045, respectively, which were not significantly different from zero ($p > 0.24$). In contrast, monkey G's sample had more near neurons, with a median *SDPI* of 0.016, which was significantly positive ($p = 0.025$).

Figure 5 shows the relationship between the *TDBI* and *SDPI*. For all three animals, the *TDBI* was significantly correlated with the *SDPI* (Pearson correlation coefficient $r = 0.57$, 0.31, and 0.24 for monkeys B, G, and R, respectively, $p < 0.01$). The positive correlation was robust regardless of whether monkeys B and G were cued to attend the near surface or the far surface (Fig. 5A, B), or whether attention was directed away from the RFs (Fig. 5C). For monkey G, even though most of the neurons showed a bias toward the near disparity of the two surfaces, the positive correlation between the *TDBI* and *SDPI* still holds. These results suggest a common trend across the animals that the disparity bias of a neuron in response to two surfaces is positively correlated with the disparity preference of the neuron to a single surface.

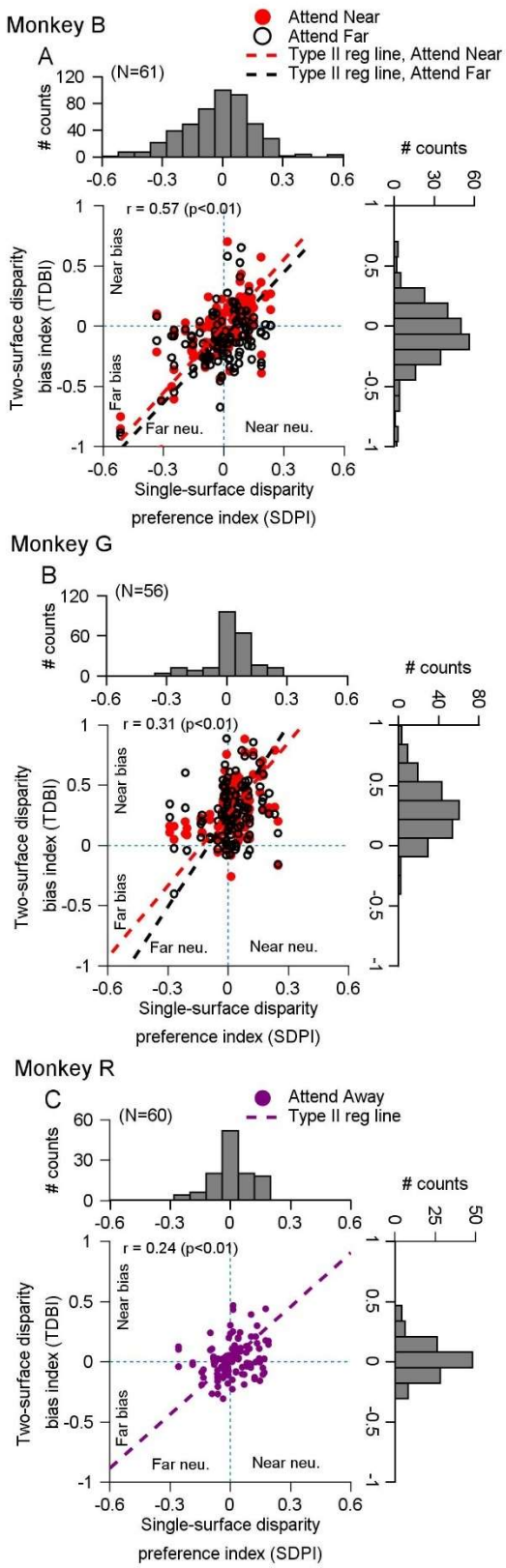


Figure 5. Relationship between the disparity bias to two surfaces and the disparity preference for a single surface. A. Scatter plot showing two-surface disparity bias index (*TDBI*) versus single-surface disparity preference index (*SDPI*) for each neuron from monkey B. Solid red circles represent attending near conditions, and open black circles represent attending far conditions. Each neuron contributes four data points (60° and 120° direction separations under both attention conditions). Dashed lines show Type II regression fits for attending near (red) and far (black) conditions. The Pearson correlation coefficient (*r*) is based on pooled results. Histograms display the distributions of *TDBI* (right) and *SDPI* (top). **B.** Results from monkey G. **C.** Results from monkey R, with attention directed away from the RFs.

Timecourse of the neuronal response tuning to bi-directional stimuli moving at different depths

To better understand the nature of disparity bias and its underlying mechanisms, we analyzed the timecourse of response tuning curves to bi-directional stimuli at different depths. We aimed to address three key questions. First, does disparity bias emerge immediately at the start of the neural response or develop over time? Second, what is the relationship between disparity bias and attention modulation? Third, across neurons preferring near and far disparities, does the MT neuron population exhibit an overall bias towards the near or far disparity of the two surfaces?

Figure 6 shows the population-averaged tuning curves of monkey B to the bi-directional stimuli, calculated within a time window of 50 ms and sliding at a step of 10 ms. As in Figure 3, we combined the conditions in which the near component moved at either the clockwise or the counterclockwise side of the two directions and plotted the near component at the clockwise side of the two motion directions. As such, at positive VA directions, the near component moved in a more preferred direction of the neurons and elicited a stronger response than the far component (Fig. 6A1-D1). Specifically, at the VA direction of 30º and 60º, the near component moved in the PD (0º) of the neuron when the direction separation was 60º and 120º, respectively. We referred to VA 30º or 60º as "Near PD" for direction separation of 60º and 120º, respectively. Conversely, at negative VA directions, the far component moved in a more preferred direction and elicited a stronger response (Figs. 6A1-D1). At the VA direction of -30º and -60º, the far component moved in the PD (0º) at the direction separation of 60º and 120º, respectively. We referred to VA -30º or -60º as "Far PD". When the response tuning to the bi-directional stimuli is biased toward the near component, the tuning curve should skew to the positive VA directions. Likewise, when the bias is toward the far component, the tuning should skew to the negative VA directions.

At 60º DS, the near neurons showed a near bias slightly delayed for about 10-20 ms relative to the neural response (Fig. 6A2, A3, Supplementary Fig. 1A). Attending the near surface made the near bias stronger and more persistent over time (Fig. 6A2), in comparison to attending the far surface (Fig. 6A3). Attending the far surface did not abolish the initial near bias, but later shifted the tuning back and forth between the near and far components and made the tuning broader than attending near (Fig. 6A3). The far neurons showed a far bias delayed about 40 ms relative to the neural response (Fig. 6B2, B3, Supplementary Figure SFig. 1C). The far bias of the far neurons appeared later than the near bias of the near neurons. Attending the far surface made the far bias more robust and persistent over time (Fig. 6B3) than attending near (Fig. 6B2). For both the near and far neurons, the average of the responses to the near and far components (*Ravg*) was roughly balanced around VA 0º, with a slight near (Fig. 6A4) and far (Fig. 6B4) skewness, respectively, due to the difference in the responses to the near and far components.

Similarly, at 120º DS, the near neurons showed a near bias (Fig. 6C2, C3), and the far neurons showed a far bias (Fig. 6D2, D3). The disparity bias appeared sooner than that at 60º DS and (SFig. 1B vs. 1A, 1D vs. 1C), again, the near bias of the near neurons emerged sooner than the far bias of the far neurons (SFig. 1B vs. 1D). Selective attention modulated the responses but was not sufficient to change the disparity bias according to attention (Fig. 6C2, C3, D2, D3). For both the near and far neurons, the *Ravg* was slightly skewed toward the neurons' preferred disparity (Fig. 6C4, 6D4).

To separate the disparity bias from the effect of selective attention, we averaged the tuning curves obtained from "Attend Near" and "Attend Far" conditions. If the effects of attending to the near and far surfaces on the direction tuning curve of the same neuron have similar magnitude and timecourse, averaging the two would neutralize the impact of attention modulation. Consequently, any remaining bias in the tuning curve would be attributed to disparity bias. To illustrate how the recorded response tuning to the bi-directional stimuli differed from the average of the component responses over time, we subtracted *Ravg* from the attention-averaged responses (Fig. 6A5-D5).

At both 60º and 120º DS, we found that the near neurons showed a sooner, initially stronger and more transient near bias (Fig. 6A5, C5) than the far bias of the far neurons (Fig. 6B5, D5). Furthermore, the near bias at 60º DS was slightly delayed relative to that at 120º DS (comparing Fig. 6A5 with C5, and Fig. 6B5 with D5, SFig. 1A with 1B). Relative to response averaging, at 60º DS, the disparity bias was reflected in response facilitation at the biased side, and suppression near VA 0º and the other side of the VA directions (Fig. 6A5, B5). At 120º DS, the disparity bias was reflected in stronger response facilitation on the biased side than on the other side (Fig. 6C5, D5). The suppression near VA 0º was stronger for the near neurons than the far neurons (comparing Fig. 6A5 vs B5, and C5 vs D5). Interestingly, at 60º DS, the far neurons initially showed a near bias, relative to the average of the component responses, for about 40 ms before changing to the far bias (Fig. 6B5 and the thin purple curve in SFig. 1C).

To examine the effects of selective attention, we subtracted the response tuning of the attending-far condition from the attending-near condition for the near neurons (Fig. 6A6, C6), and subtracted the attending-near condition from the attending-far condition for the far neurons (Fig. 6B6, D6). For near neurons, the attention modulation was reflected in response facilitation at positive VA directions around "Near PD" and/or suppression at negative VA directions around "Far PD" (Fig. 6A6, C6). The facilitation (Fig. 6A6) and suppression (Fig. 6A6, C6) occurred soon after the response onset. In contrast, for far neurons, the attention modulation was mainly reflected in much-delayed facilitation at negative VA directions around "Far PD", without suppression at the positive VA directions around "Near PD" (Fig. 6B6, D6).

In summary, the timecourse analysis revealed several findings from monkey B. First, the disparity bias is often delayed relative to the response onset. The delay is slightly longer at the smaller (60º) than larger (120º) DS, likely reflecting a longer processing time needed for segmentation at smaller DS. Second, the near bias of the near neuron emerges sooner than the far bias of the far neuron. Third, the disparity bias is not the result of attention modulation but attention can modulate the disparity bias. Attention modulation for the near neurons also occurs sooner than the far neurons (comparing Fig. 6A6, C6 with B6, D6).

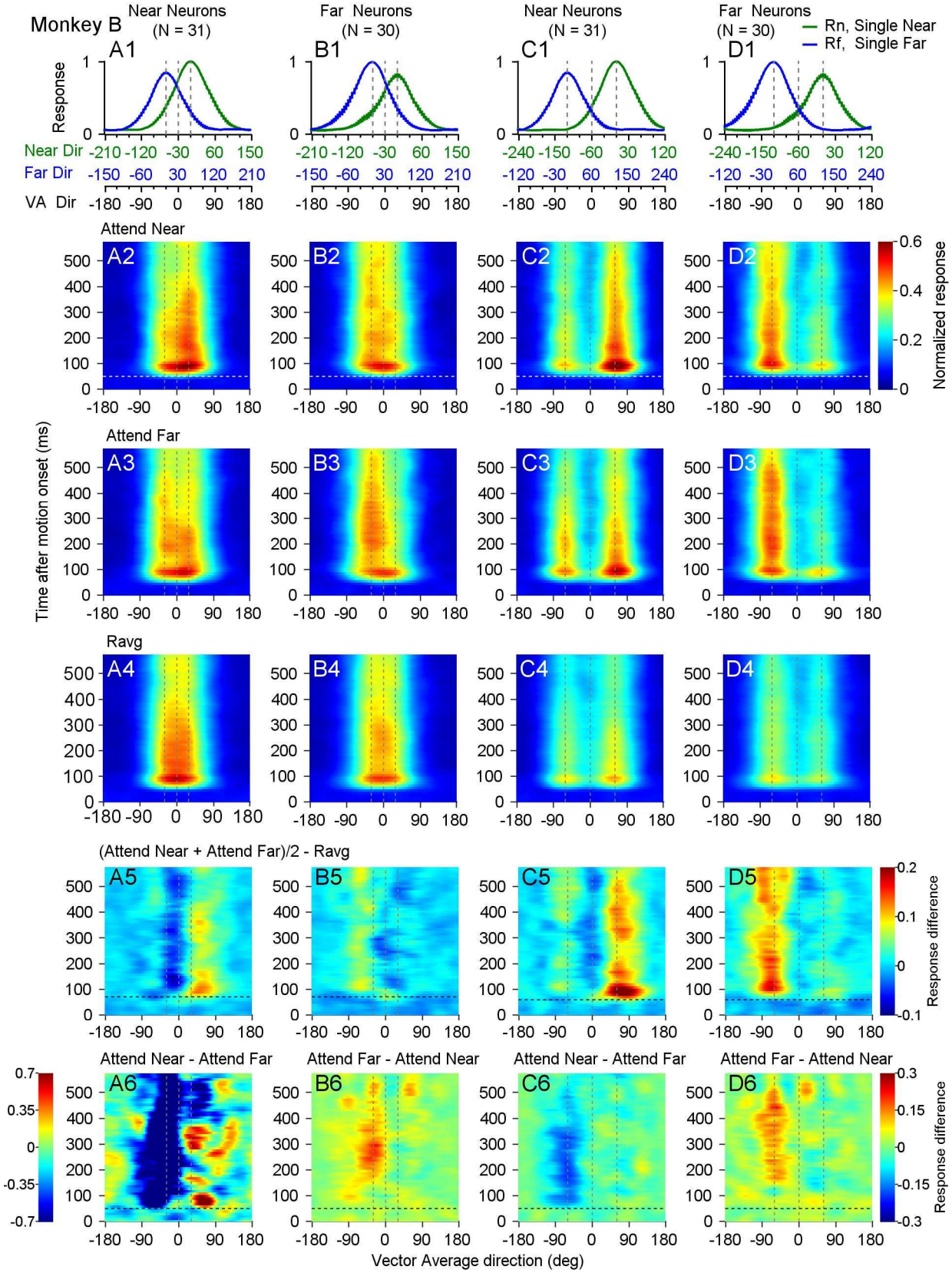

Monkey B
Near Neurons (N = 31)
Far Neurons (N = 30)
Near Neurons (N = 31)
Far Neurons (N = 30)
Rn, Single Near
Rf, Single Far
A1
B1
C1
D1
Response
Near Dir
Far Dir
VA Dir
Attend Near
Attend Far
Ravg
(Attend Near + Attend Far)/2 - Ravg
Attend Near - Attend Far
Attend Far - Attend Near
Time after motion onset (ms)
Normalized response
Response difference
Vector Average direction (deg)

Figure 6. Timecourse of response tuning to bi-directional stimuli moving at different depths in monkey B. Columns A, B. Responses to 60° direction separation. **Columns C, D.** Responses to 120° direction separation. **Columns A, C.** Near neurons. **Columns B, D.** Far neurons. **A1-D1.** Population-averaged response tuning curves to single component directions at near (green) or far (blue) depth. The width of the curve for R_n, R_f represents the standard error. Vertical dashed lines mark the VA direction at the neuron's PD (0°) (middle line), and when far (left line) or near (right line) components move in the PD. **A2-D2.** Timecourse of population-averaged tuning curves to bi-directional stimuli under Attend Near condition. Horizontal dashed lines at 50 ms after motion onset roughly mark the response onset. **A3-D3.** Responses under Attend Far condition. **A4-D4.** Average responses to near and far components (*Ravg*). **A5-D5.** Average responses under Attend Near and Attend Far conditions, subtracting *Ravg*. Horizontal dashed lines at 70 ms (A5) roughly indicate the onset of near bias at 60° direction separation. Horizontal dashed lines at 60 ms (C5) roughly indicate the onset of near bias at 120° direction separation. The far bias of far neurons is later than the near bias of near neurons (B5, D5). **A6-D6.** Difference between responses under two attention conditions. A6, C6. Attend Near – Attend Far. B6, D6. Attend Far – Attend Near. Horizontal dashed lines at 50 ms (A6-D6) roughly indicate the onset of attention effect for near neurons (A6, C6). The effect of attention for far neurons is later than for near neurons (B6, D6).

Figure 7 shows the timecourse of direction tuning from monkey G. At 60º DS, the near neurons showed the development of near bias over a period of about 80 ms, starting from around 110 ms to fully biased to the near side around 190 ms (Fig. 7A2, A3, A5, SFig. 2A). At 120º DS, the near neurons showed a stronger near bias than at 60º DS (Fig. 7C2, C3, C5, SFig. 2B). The near bias was stronger and more persistent when attention was directed to the near surface (Fig. 7A2, C2) than directed to the far surface (Fig. 7A3, C3). Far neurons of monkey G also showed a near bias (Fig. 7B2, B3, B5, D2, D3, D5, SFig. 2C, 2D), displaying a timecourse similar to but slightly later than that of the near neurons (SFig. 2). The responses of the far neurons at the positive VA directions around the "near PD" were robust and sustained throughout the motion period, making the tuning curve broader than the near neurons (comparing Fig. 7B2 with A2, and B3 with A3).

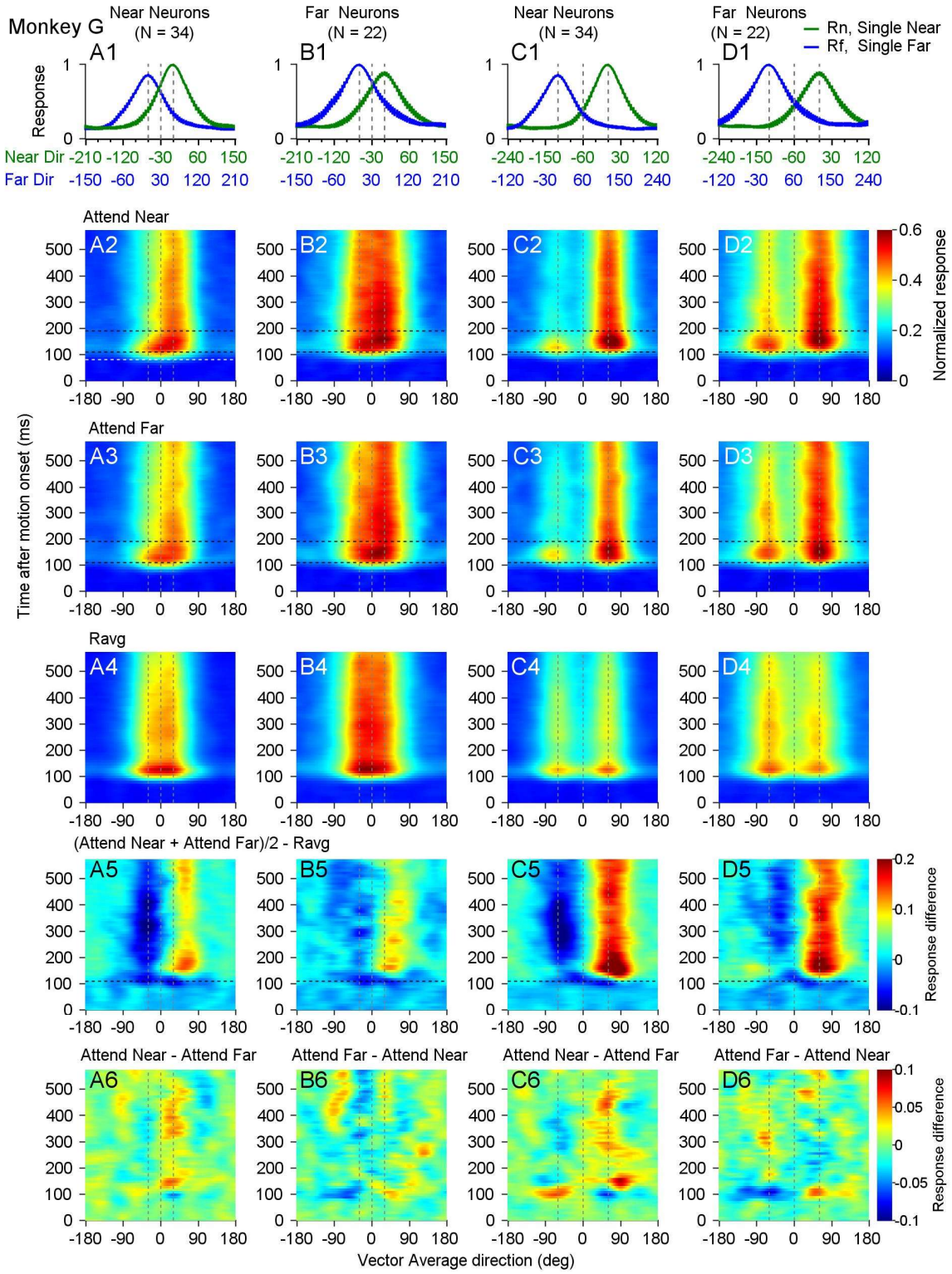
Although the weight analysis based on the full tuning curve and the entire motion response period did not reveal a significant attention effect for monkey G (Fig. 4B), the timecourse analysis indicated a small effect of attention modulation, which was more evident in the near neurons than in the far neurons. The attention modulation for the near neurons mainly reflected in response facilitation at the positive VA directions around the "near PD" when the response of attending far was subtracted from attending near (Fig. 7A6, C6). For near neurons, there was small and brief facilitation toward negative VA directions around the far PD at the beginning of neural response (about 100 ms after motion onset) (Fig. 7A6, C6), suggesting that, in the attending-near condition, monkey G may have initially attended to the far surface briefly before attending to the cued near surface.

Far neurons showed little attention modulation at 60º DS (Fig. 7B6), but showed small and delayed facilitation at the negative VA directions around the "far PD" at 120º DS when the response tuning of attending near was subtracted from attending far (Fig. 7D6). Similar to the near neurons, the far neurons showed a small initial facilitation toward positive VA directions around the near PD at about 100 ms (Fig. 7B6, D6), suggesting that, in the attending-far condition,

monkey G may have attended to the near surface briefly before attending to the cued far surface. Nevertheless, regardless of the effect of attention modulation, the near bias was robust throughout the motion response period in both the near and far neurons (Fig. 7A2-D2, A3-D3, A5-D5).

Figure 8 shows the timecourse of the response tuning to bi-directional stimuli at different depths in monkey R, with the animal's attention directed away from the RFs. The near and far neurons showed disparity bias toward their preferred disparity (Fig. 8A2-D2, A4-D4). The near bias of the near neurons appeared sooner than the far bias of the far neurons (Fig. 8A4-D4, SFig. 3). For the far neurons, the early responses were roughly balanced between VA directions around the far PD and near PD, and the far bias only became evident 160 ms and 100 ms after motion onset for DS60º and 120º, respectively (Fig. 8B4, D4, SFig. 3C, D). As found in monkey B, the far neurons from monkey R initially showed a near bias relative to the average of the component responses, before gradually changing to the far bias (thin purple curves in SFig. 3C, 3D).

Figure 7. Timecourse of response tuning to bi-directional stimuli moving at different depths in monkey G. The convention follows that of Figure 6. **A1-D1.** Population-averaged response tuning curves to single component directions at near (green) or far (blue) depth. **A2-D2.** Timecourse of population-averaged tuning curves to bi-directional stimuli under Attend Near condition. In A2, the white horizontal dashed line at 80 ms indicates response onset and the two black dashed lines at 110 ms and 190 ms indicate the time period during which the near bias of near neurons develops. **A3-D3.** Responses under Attend Far condition. The black dashed lines at 110 ms and 190 ms are also plotted in B2-D2 and A3-D3 for reference. **A4-D4.** Average responses to near and far components (*Ravg*). **A5-D5.** Average responses under Attend Near and Attend Far conditions, subtracting *Ravg*. The black dashed lines are at 110 ms as a reference. **A6-D6.** Difference between responses under two attention conditions. Vertical dashed lines in all panels mark the VA direction at the neuron's PD (0°) (middle line), the far PD (left line), and the near PD (right line).


Monkey G
Near Neurons (N = 34)
Far Neurons (N = 22)
Near Neurons (N = 34)
Far Neurons (N = 22)
Rn, Single Near
Rf, Single Far
A1
B1
C1
D1
Response
Near Dir -210 -120 -30 60 150
Far Dir -150 -60 30 120 210
-210 -120 -30 60 150
-150 -60 30 120 210
-240 -150 -60 30 120
-120 -30 60 150 240
-240 -150 -60 30 120
-120 -30 60 150 240
Attend Near
Attend Far
Ravg
(Attend Near + Attend Far)/2 - Ravg
Attend Near - Attend Far
Attend Far - Attend Near
Attend Near - Attend Far
Attend Far - Attend Near
Time after motion onset (ms)
Normalized response
Response difference
Vector Average direction (deg)

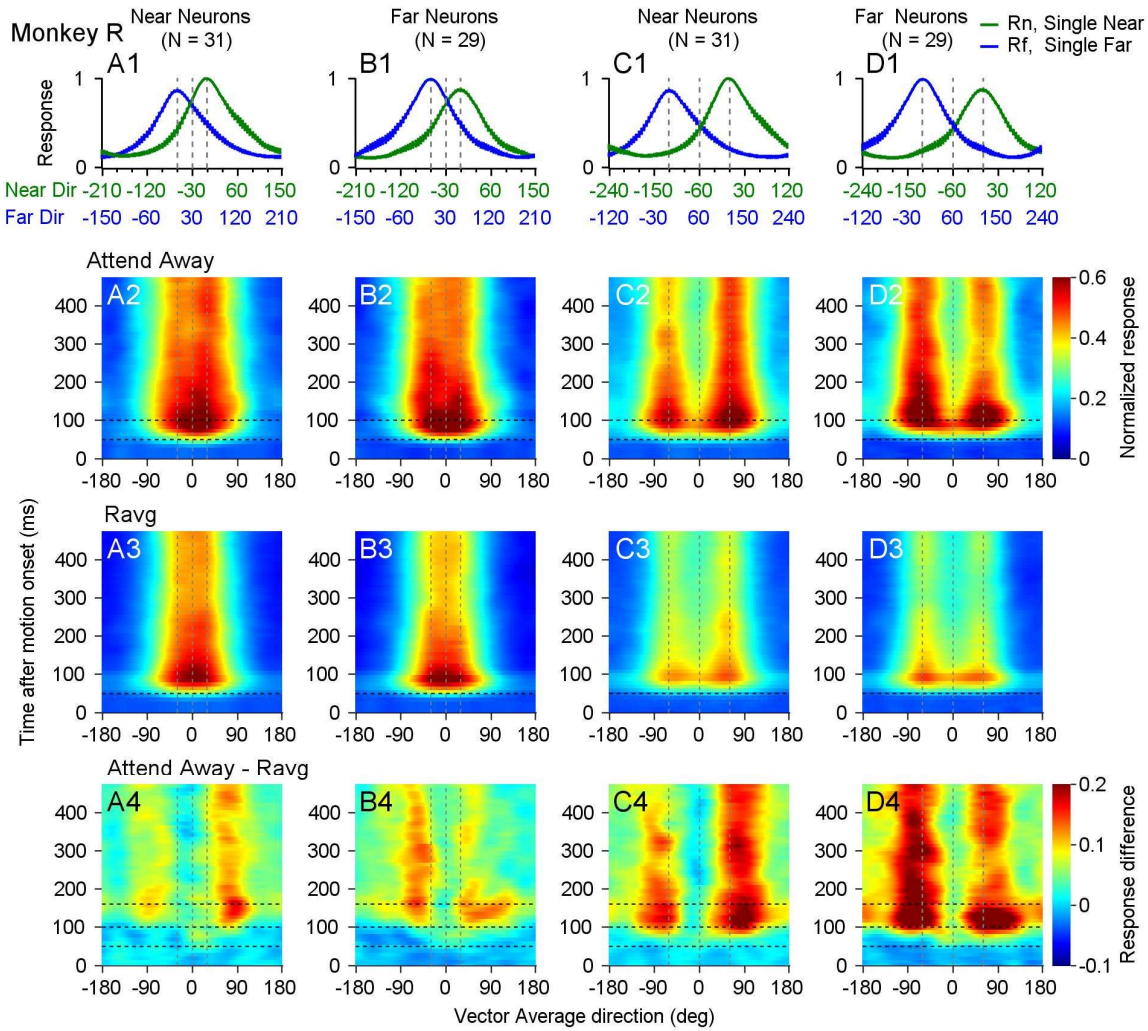


Figure 8. Timecourse of response tuning to bi-directional stimuli moving at different depths in monkey R. The convention is similar to that of Figures 6 and 7. **A1-D1.** Population-averaged response tuning curves to single component directions at near (green) or far (blue) depth. **A2-D2.** Timecourse of population-averaged tuning curves to bi-directional stimuli with the animal's attention directed away from the RFs. The two horizontal dashed lines are at 50 ms and 100 ms after motion onset as refenerences. **A3-D3.** Average responses to near and far components (*Ravg*). The horizontal dashed line is at 50 ms. **A4-D4.** Responses to bi-directional stimuli subtracting *Ravg*. The three horizontal dashed lines are at 50 ms, 100 ms, and 160 ms after motion onset. Vertical dashed lines in all panels mark the VA PD (middle line), the far PD (left line), and the near PD (right line).

So far, we have characterized the timecourse of disparity bias separately for near and far neurons. Next, we examined whether MT neurons as a whole showed an overall bias toward the near or far disparity of overlapping surfaces over time. We averaged the tuning curves to the bi-directional stimuli within a sliding time window of 50 ms across near and far neurons for each animal. To assess disparity bias, we subtracted *Ravg* from the average Attend-Near and Attend-

Far responses for monkeys B (Fig. 9A1, B1) and G (Fig. 9A2, B2), and from the attention-away responses for monkey R (Fig. 9A3, B3). During the early response, all three animals exhibited a near bias for both 60º DS (Fig. 9A1-A3) and 120º DS (Fig. 9B1-B3), with stronger responses at positive VA directions than at negative VA directions.

To quantify the disparity bias, we flipped the responses at negative VA directions relative to VA 0° and subtracted them from the responses at positive VA directions (Fig. 9C1-C3 and D1-D3). Positive values of the response difference (shown in black-red-yellow) indicate a near bias, while negative values (shown in blue-cyan) indicate a far bias. For all three animals, the response differences in the early time windows (integrated within VA directions ranging from 15º to 120º in Fig. 9C1-C3, D1-D3) were significantly positive ($p < 0.01$ for monkeys B and G, and $p \leq 0.05$ for monkey R, signed-rank test), indicating a near bias (Table 1).

Table 1. Time period showing significant near bias

Time after motion onset	Mid-point of 50-ms time bins (Time period measured in 10-ms bin) (p value)	
Direction separation	**DS 60°**	**DS 120°**
Monkey B (N = 61)	55 – 85 ms (30 – 110 ms) ($p < 0.01$)	45 – 85 ms (20 – 110 ms) ($p < 0.01$)
Monkey G (N = 56)	115 – 575 ms (90 – 600 ms) ($p < 0.01$)	105 – 575 ms (80 – 600 ms) ($p < 0.01$)
Monkey R (N = 60)	35 – 55 ms (10 – 80 ms) ($p < 0.05$)	75 ms (50 – 100 ms) ($p = 0.05$)

Whether the population-averaged response tuning showed a near or a far bias depended on the proportion of near and far neurons in the data sample and their disparity preferences as measured by *SDPI*. As shown in Figure 5, data samples from monkeys B and R were roughly balanced between near and far neurons. Therefore, the early near bias observed in the population-averaged tuning curves of monkeys B and R (Fig. 9C1, D1, C3, D3) cannot be attributed to the incidental disparity preferences of the neurons in the data samples. Our findings indicate that near neurons showed an earlier near bias than far neurons showing far bias in monkeys B (Fig. 6, SFig. 1) and R (Fig. 8, SFig. 3), likely contributing to the early near bias observed in the overall population. While monkey G's sample had more near neurons than far neurons, the consistent near bias found in both early and late responses (Fig. 9C2, D2) is due to both near and far neurons exhibiting the near bias (Fig. 7, SFig. 2).

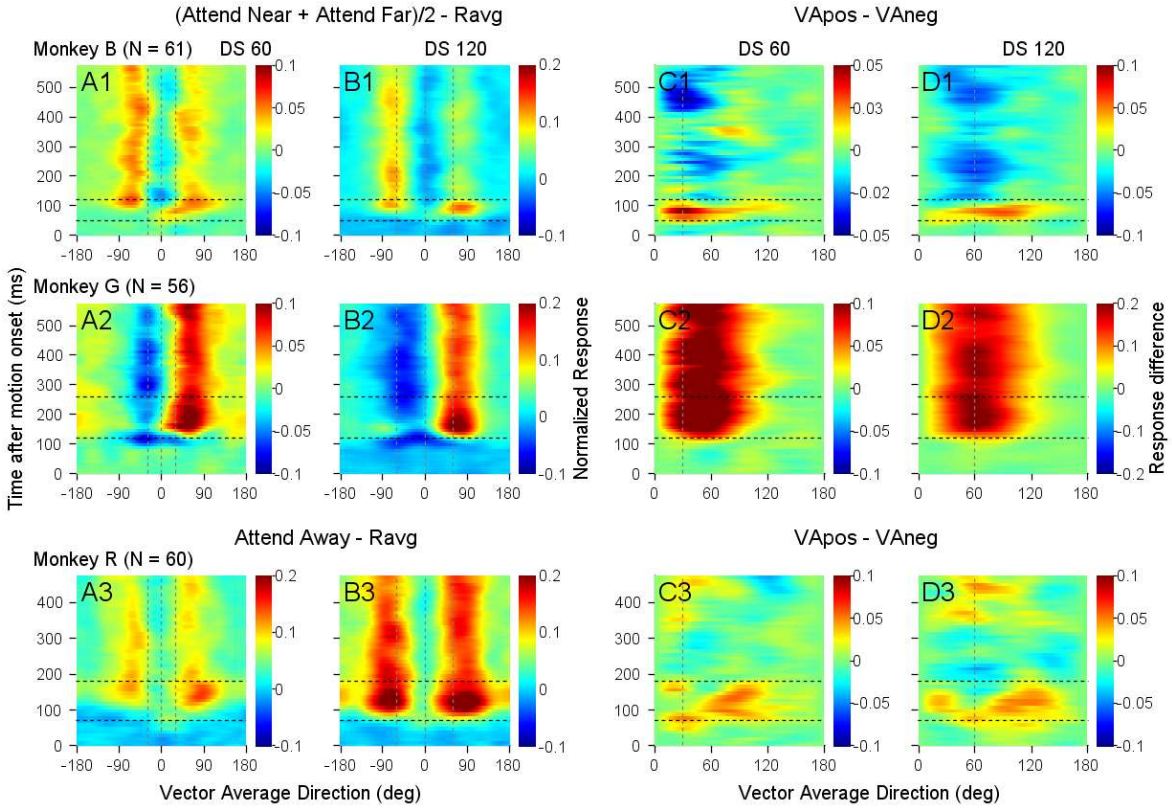


Figure 9. Early near bias across all neurons in three monkeys. A1-D1. Responses from monkey B. Horizontal dashed lines indicate 50 and 120 ms after motion onset. **A2-D2.** Responses from monkey G. Horizontal lines indicate 120 and 260 ms. **A3-D3.** Responses from monkey R. Horizontal lines indicate 70 and 180 ms. **Columns A, C.** Responses to 60° DS. **Columns B, D.** Responses to 120° DS. **A1, B1, A2, B2.** Average responses under Attend Near and Attend Far conditions, subtracting Ravg. **A3, B3.** Responses to bi-directional stimuli with attention directed away from the RFs, subtracting Ravg. Vertical dashed lines in columns A and B mark VA PD (middle line), far PD (left line), and near PD (right line). **Columns C, D.** Responses at negative VA directions were flipped relative to VA 0° and subtracted from responses at positive VA directions. The vertical dashed line indicates the component PD.

Attention modulation of neuronal response is object-based

Figure 4 describes the effect of attention on the full response tuning curve to bi-directional stimuli, as measured by the impact on the weights for the near and far components. We further asked whether the effect of attention was feature-based or object-based when the animals were cued to attend a stereoscopic depth of a surface. If the attention was feature-based (disparity-based), we expected to find that when attention was directed to a more preferred disparity of a neuron, the neuronal response would be enhanced regardless of the motion direction of the surface. On the other hand, if attention was object-based (surface-based), we expected to find that when attention was directed to a surface that elicited a stronger response than the other surface, the neuronal response would be enhanced regardless of whether the attended surface was located at the more or less preferred disparity of the neuron.

To answer this question, we measured the effect of attention directed to a stimulus component that was located at either the more preferred or less preferred disparity and moving in either the more preferred or less preferred direction of the neuron. We quantified the effect using two attention modulation indices (AMI). One index *AMI_disparity* would be positive across conditions if the effect of attention was feature-based (Eq. 5). Whereas the other index *AMI_surface* would be positive if the effect of attention was object-based (Eq. 6).

$$AMI_disparity = \frac{R_{att_PrefDsp} - R_{att_nPrefDsp}}{R_{att_PrefDsp} + R_{att_nPrefDsp}} \quad , \qquad (5)$$

in which $R_{att_PrefDsp}$ and $R_{att_nPrefDsp}$ were the responses to the overlapping stimuli when attention was directed to the more preferred disparity and less preferred disparity of the neuron, respectively.

$$AMI_surface = \frac{R_{att_strongerSurf} - R_{att_weakerSurf}}{R_{att_strongerSurf} + R_{att_weakerSurf}} \quad , \qquad (6)$$

in which $R_{att_strongerSurf}$ and $R_{att_weakerSurf}$ were the responses to the overlapping stimuli when attention was directed to a surface that elicited a stronger response and a weaker response of the neuron, respectively.

We found that the results from monkey B were consistent with object-based attention. Figure 10A and B showed the response tuning curves of two representative neurons from monkey B, one neuron preferred the near disparity (Fig. 10A) and the other preferred the far disparity (Fig. 10B). When attention was directed to a surface that elicited a stronger component response, regardless of whether the surface was located at the more preferred or less preferred disparity of the neuron, the response was always enhanced relative to attending the other surface (Fig. 10A, 10B). This trend was consistent across the neuronal population as captured by always positive *AMI_surface* values (Fig. 10C) for the near and far neurons, and for 60° and 120° DSs, respectively.

In contrast, when attention was directed to the surface located at the more preferred disparity of the neuron (near in Fig. 10A and far in 10B) and the surface moved in a less preferred direction, resulting in a weaker component response, the response to both surfaces was reduced relative to attending to the other surface (the red curve at "Far PD" in Fig. 10A and the black curve at "Near PD" in Fig. 10B). Furthermore, when attention was directed to the surface located at the less preferred disparity of the neuron (far in Fig. 10A and near in 10B) and the surface moved in a more preferred direction, resulting in a stronger component response, the response to both surfaces was nevertheless enhanced (the black curve at "far PD" in Fig. 10A and the red curve at "near PD" in Fig. 10B). In other words, the enhancement or reduction of the response by attention was not dependent on whether the attended surface was the more preferred or less preferred disparity of the neuron, respectively, but rather dependent on whether the attended surface elicted a stronger or weaker component response. As a result, *AMI_disparity* averaged across the neuronal population changed sign with stimulus conditions (Fig. 10D), suggesting that the effect of attention was not feature-based.

The attention effect found in monkey G was minimal when measured by the response weights based on the response during the whole stimulus motion period (Fig. 4B). However, the timecourse of the neuronal responses revealed that the attention effect in monkey G was visible in most conditions between 110 ms and 410 ms after motion onset (Fig. 7A6, C6, D6), except for far neurons responding to a 60º DS (Fig. 7B6). Therefore, we used this time window to calculate direction tuning curves (Fig. 10E, F) and the *AMIs*. Although the attention effect in monkey G was not statistically significant for most conditions, except for the responses of the far neurons to 60º DS, the trend of *AMI* in monkey G (Fig. 10G, H) was consistent with monkey B (Fig. 10C, D) and supported the idea of object-based attention modulation.

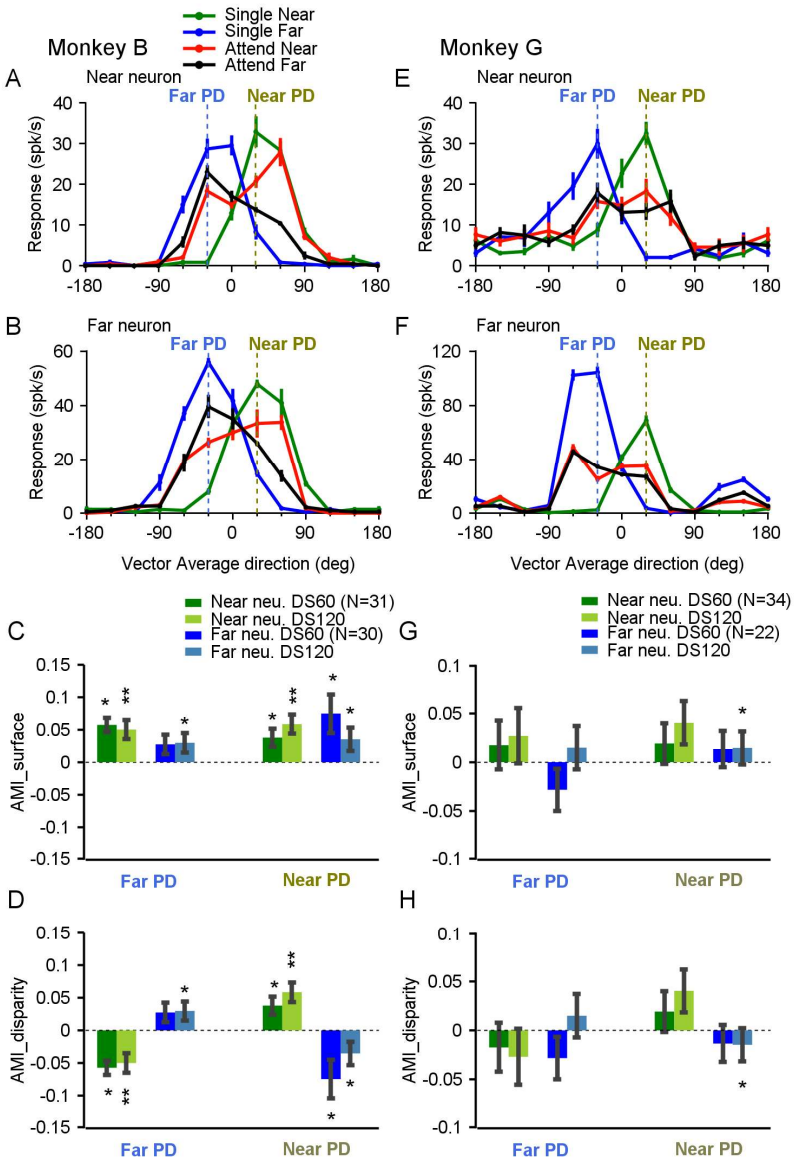


Figure 10. Attention modulation of neuronal responses to bi-directional stimuli moving at two depths for monkey B (A-D) and monkey G (E-H). **A, B, E, F.** Tuning curves of four example neurons to bi-directional stimuli. The convention follows that of Figure 2. Error bars represent the standard error. The vertical dashed lines indicate the Near PD and the Far PD. **C, G.** Population-averaged object-based attention modulation indices (*AMI_surface*) were calculated separately using responses to bi-directional stimuli at the Far PD and the Near PD, across near neurons and far neurons, and for 60° and 120° DSs. **D, H.** Disparity-based attention modulation indices (*AMI_disparity*). Responses of monkey B were calculated from 0 to 600 ms following the motion onset. Responses of monkey G were calculated from 110 to 410 ms. (**) indicates $p < 0.01$, (*) indicates $p < 0.05$, and no asterisk indicates $p > 0.05$ (one-sided signed-rank test to compare *AMI* with zero).

Modeling neural responses to two surfaces at different depths using a unified model

Our neural data revealed that MT responses to overlapping moving stimuli separated in depth can be explained by a weighted sum of responses to individual stimulus components presented alone. These weights were influenced by the neuron's disparity preference to single surfaces (Fig. 5) and the effects of selective attention (Fig. 4 and Fig. 10). While the general trend was consistent across animals, we also observed individual differences. Specifically, in monkey G, both near and far neurons exhibited a dominant near bias, though near neurons showed a stronger bias than far neurons (Fig. 5). In contrast, in monkey B, the near neurons showed a near bias and the far neurons showed a far bias. These findings prompted us to seek a unified model capable of explaining the seemingly complex relationships.

The divisive normalization model (Carandini and Heeger, 2012) has been used to explain a wide array of phenomena, including neuronal responses to multiple visual stimuli (Britten and Heuer, 1999; Heuer and Britten, 2002; Busse et al., 2009; Xiao et al., 2014; Xiao and Huang, 2015; Bao and Tsao, 2018; Wiesner et al., 2020; Huang et al., 2023) and the effects of attention (Reynolds and Heeger, 2009; Lee and Maunsell, 2009; Ni et al., 2012; Ni and Maunsell, 2019). While in the model, the division by the activity of a population of neurons (i.e., normalization) is well established, the factors contributing to the numerator are less clear. Typically, in response to multiple stimuli, the signal strength (such as luminance contrast or motion coherence) of each stimulus component is used in the numerator (e.g., Busse et al., 2009; Xiao et al., 2014). However, defining the signal strength of a sensory stimulus is not always straightforward (Wiesner et al., 2020). We have proposed that the weighting to a stimulus component of multiple stimuli is proportional to the activity of a population of neurons elicited by the same stimulus component (Xiao et al., 2014; Wiesner et al., 2020; Huang et al., 2023). We refer to this neuron population as the “weighting pool”. The nature and the scope of the "weighting pool" remain unclear.

Here, we expand on this idea and suggest that the size of the weighting pool can be variable. At one extreme, the weighting pool may encompass neurons with RFs at the same location but with all possible disparity preferences, a scenario we call “global pooling”. Conversely, at the other extreme, the weighting pool may consist solely of neurons with the same preferred disparity as the neuron under study, which we term "local pooling". The size of the weighting pool can vary between these two extremes. Equations 7-10 describe the model without considering the effect of attention.

$$R_k(\theta_n, \theta_f) = \frac{S_k(n, d_{pref})^m}{S_k(n, d_{pref})^m + \alpha \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_n) + \frac{S_k(f, d_{pref})^m}{S_k(n, d_{pref})^m + \alpha \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_f) + c \,, \quad (7)$$

in which $R_k(\theta_n, \theta_f)$ is the model-predicted response of neuron k to two motion directions θ_n and θ_f, moving at the near (n) and the far (f) disparity. $R_k(\theta_n)$ and $R_k(\theta_f)$ are the measured responses of the same neuron to the individual components θ_n and θ_f, respectively. $m, \alpha, \sigma,$ and c are model parameters. S_k is the summed response of the weighting pool associated with neuron k, which has a preferred disparity d_{pref}, to the disparity of a stimulus component. In our model, S_k is calculated by taking the dot product of a Gaussian window $\boldsymbol{G}$, centered on d_{pref}, and the response of a population of neurons with all possible disparity preferences to the disparity of a stimulus component (Eq. 8).

$$S_k(d_{stim},\ d_{pref}) = \boldsymbol{G}(d_{pref}, v) \bullet \boldsymbol{pR}(d_{stim}) \,, \quad (8)$$

in which d_{stim} is the disparity of a stimulus component, $\boldsymbol{pR}$ is the response of a population of neurons with all possible disparity preferences (Eq. 9), and $\boldsymbol{G}$ is a Gaussian function with a standard deviation of v and centered on d_{pref} (Eq. 10).

$$\boldsymbol{pR}\ (d_{stim}) = [f_i(d_{stim},\ \ d_{pref,i})], \quad i = 1 \text{ to N} \,, \quad (9)$$

in which f_i is the disparity tuning function of neuron i in response to single disparities. $d_{pref,i}$ is the preferred disparity of neuron i.

$$\boldsymbol{G}(d_{pref}, v) = [\frac{1}{\sqrt{2\pi v}} e^{\frac{-(d-d_{pref})^2}{2v^2}}] \,, \quad (10)$$

$$d \in [-1.5,\ 1.5] \text{ and matches } d_{pref,i}\ (i = 1 \text{ to N}) \text{ of } \boldsymbol{pR}.$$

In other words, the extent of the weighting pool is controlled by the width of the Gaussian window $\boldsymbol{G}$. The parameter v is the standard deviation of $\boldsymbol{G}$. Both $\boldsymbol{pR}$ and $\boldsymbol{G}$ are vectors of the same length N. Given the relatively small number of neurons in our data set, the proportion of near and far neurons may not accurately reflect the true distribution of disparity preferences among MT neurons. Therefore, we used the disparity tuning data from a large sample of MT neurons (N = 479) as reported by DeAngelis and Uka (2003) to estimate $\boldsymbol{pR}$.

To model the effect of attention, we added the term β to equation 7. The model-predicted responses $R_k(\theta_{na}, \theta_f)$ and $R_k(\theta_n, \theta_{fa})$ correspond to attention being directed to the near-surface and the far-surface, respectively (Eqs. 11, 12). The values of β_{An} and β_{Af} were derived from the weights of a linear weighted summation (LWS) model fit (see Methods and Eq. 16).

$$R_k(\theta_{na}, \theta_f) = \frac{\beta_{An} \bullet S_k(n, d_{pref})^m}{\beta_{An} \bullet S_k(n, d_{pref})^m + \alpha \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_n) + \frac{S_k(f, d_{pref})^m}{\beta_{An} \bullet S_k(n, d_{pref})^m + \alpha \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_f) + c \,, \quad (11)$$

$$R_k(\theta_n, \theta_{fa}) = \frac{S_k(n, d_{pref})^m}{S_k(n, d_{pref})^m + \alpha \bullet \beta_{Af} \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_n) + \frac{\beta_{Af} \bullet S_k(f, d_{pref})^m}{S_k(n, d_{pref})^m + \alpha \bullet \beta_{Af} \bullet S_k(f, d_{pref})^m + \sigma} R_k(\theta_f) + c \,, \quad (12)$$

The ranges of the model parameters were constrained as follows: $0 < m < 100$, $c > 0$, $\sigma > 0$, and $\alpha > 0$. The parameter α scales the relative contributions of $S_k(n)$ and $S_k(f)$ to the denominator as in tuned normalization (Ni et al., 2012; Rust et al., 2006; Carandini et al., 1997).

Figure 11A illustrates the population neural response $\boldsymbol{pR}$ to a single near disparity of -0.1º and a far disparity of 0.1º of our visual stimuli, with the single stimulus moving in the preferred directions of the neurons. Figure 11A also displays examples of Gaussian function $\boldsymbol{G}$ with three different widths, centered on disparities of -0.5º and 0.5º, demonstrating the transition from local pooling to global pooling. Due to the limited range of neurons' preferred disparities in the data set, the Gaussian window is unavoidably truncated by the disparity borders. In practice, we found the area under the truncated Gaussian window, as a measure of the weighting pool size, that best fits the neuronal response using Equations 11 and 12.

The model fitted the neuronal responses of monkeys B and G to the bi-directional stimuli reasonably well. For monkey B, the mean PV was 89 % (SD = 11.2), and for monkey G, it was 80% (SD = 17.5). We compared the weighting pool size that provided the best fit for the near neurons (Fig. 11B) and the far neurons (Fig. 11C) between monkeys B and G. For the near neurons, monkey B had more neurons that were best fitted with local pooling, while monkey G had more neurons that were best fitted with global pooling. However, these differences were not statistically significant as measured by 95% confidence intervals using bootstrapping (Fig. 11B).

For the far neurons, local pooling provided the best fit for significantly more neurons in monkey B than monkey G (at a normalized area of Gaussian window less than 0.4, $p < 0.05$), whereas global pooling provided the best fit for significantly more neurons in monkey G than monkey B (at area greater than 0.9, $p < 0.05$) (Fig. 11C). For 11 out of 15 far neurons (70%) from monkey G, the optimal area of the Gaussian window was greater than 0.85, indicating that global pooling best accounted for the responses of far neurons from monkey G. In contrast, local and intermediate pooling best accounted for the results from monkey B.

An intuitive explanation for these results is as follows. Because the overall MT population has more near neurons than far neurons (DeAngelis and Uka, 2003), the global pooling of MT population responses to -0.1° near disparity results in a slightly stronger response than to 0.1° far disparity. Consequently, the weight for the near component is greater than for the far component, irrespective of the disparity preference of the neuron under study. Conversely, the local pooling of MT population responses to the component disparities aligns with the disparity preference of the neuron under study. Thus, near neurons show a near bias, and far neurons show a far bias. Taken together, our findings demonstrate that a unified model with a variable size of the weighting pool and divisive normalization can explain both the disparity bias and the individual differences between the animals.

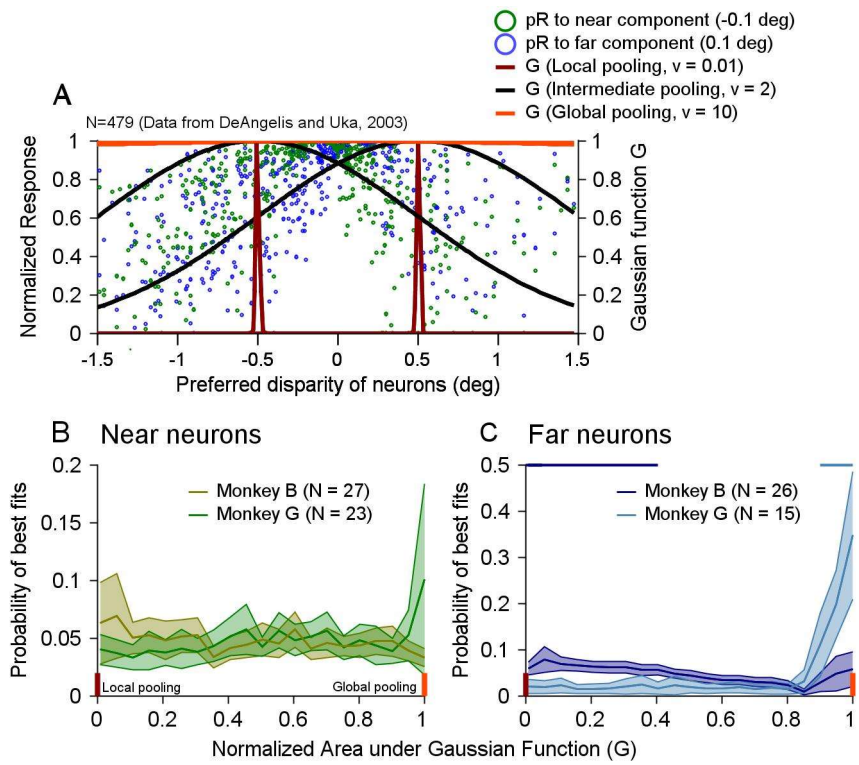


Figure 11. Modeling neural responses to bi-directional stimuli at two depths. A. MT population neural responses to component disparities of -0.1° and 0.1°, illustrating local, intermediate, and global pooling. The population responses are based on the disparity tuning curves of 479 MT neurons from DeAngelis and Uka (2003). v is the standard deviation of the Gaussian function, centered on the disparities of -0.5° and 0.5°, as two example d_{pref} of neurons under study. **B.** Distribution of pooling size providing the best fit for near neurons' responses from monkeys B and G. **C.** Distribution for far neurons, with significant differences between monkeys B and G indicated by horizontal blue and cyan lines at those pooling sizes. Error bands show 95% confidence intervals measured by bootstrapping. Short vertical brown and red bars in B and C represent local and global pooling sizes. Only neurons for which complete disparity tuning curves were obtained were included in this analysis.

Discussion

Disparity bias and implication on the rules of representing multiple visual stimuli

Our laboratory previously found that when two overlapping random-dot stimuli move in different directions at the same depth, MT neurons exhibit a robust response bias toward the stimulus with higher motion coherence or luminance contrast (Xiao et al., 2014). Nearly all MT neurons respond more strongly to higher motion coherence (Britten et al., 1992, 1993; Xiao et al., 2014) and contrast (Cheng et al., 1994; Heuer and Britten, 2002; Sclar et al., 1990), showing consistent stimulus preference at both individual and population levels. However, different neurons prefer different values for features like binocular disparity. Our current study establishes the encoding rule for representing multiple stimuli when neurons have varying preferences for a stimulus feature.

We discovered that MT neurons displayed a disparity bias toward one of two overlapping surfaces at different depths. This bias positively correlated with each neuron's preference for the disparity of a single surface when presented alone (Fig. 5). Despite individual differences among the animals, this positive correlation was consistent. Our finding may likely extend beyond binocular disparity. For instance, in another study from our lab, neuronal responses to two spatially separated stimuli moving in different directions showed a robust spatial bias, correlating with the spatial preferences of the neurons (Wiesner and Huang, 2019, SFN abstract). Thus, the disparity bias in this study may also be considered a "spatial bias," aiding in segmenting multiple visual stimuli in 3D space.

An extended normalization model and the neural mechanism underlying the disparity bias

In the standard divisive normalization model, the response to multiple stimuli is described by a weighted sum of the responses elicited by the individual stimulus components. The weight is proportional to the signal strength (such as contrast) of each stimulus component, normalized by a combination of the signal strengths of multiple stimuli (Busse et al., 2009; Carandini and Heeger, 2012; Ni et al., 2012; Xiao et al., 2014; Bao and Tsao, 2018). In terms of the neural implementation of the divisive normalization model, it has been proposed that the summed response of a population of neurons, referred to as the normalization pool, serves as the denominator in the normalization equation (Heeger, 1992; Carandini et al., 1997; Simoncelli and Heeger, 1998; Rust et al., 2006; Kouh and Poggio, 2008; Bao and Tsao, 2018; Wiesner et al., 2020). However, the neural implementation of the numerator in the normalization equation has been unclear when multiple visual stimuli are involved. We and other investigators have suggested that the summed response of a population of neurons elicited by a stimulus component may serve as the numerator and determine (together with the normalization) the weight of that component (Xiao and Huang, 2014; Bao and Tsao, 2018; Wiesner et al., 2020).

In this study, we propose a novel extension to the standard normalization model: the pool size of the neuron population in the numerator (the weighting pool) can vary. This variability allows the weighting for a stimulus component to depend on neurons with similar feature selectivity (local pooling) or those covering a full range of values (global pooling), or an intermediate range. With local pooling, a neuron's response to two stimuli would bias toward its preferred feature value, as seen in monkeys B and R. Global pooling, however, reflects the preference of a full neuron population, leading to a near bias in response to two stimuli, regardless of individual neuron preferences. In the case of binocular disparity, MT neurons exhibit a higher proportion of neurons preferring near disparity over far disparity (DeAngelis and Uka, 2003). Therefore, global pooling results in a stronger response to near disparity. When the pool size is broad enough to represent this overall near preference but not entirely invariant to individual neuron preferences, a near bias and positive correlation with the neuron’s single-stimulus disparity preference emerges, as observed in monkey G.

Our model offers a unified explanation of the disparity bias observed across different animals. In the future, it would be interesting to understand what factors (such as experience, cognition, and other individual differences) determine the size of the weighting pool. It is also plausible that pooling sizes vary for different stimulus features. For instance, recent evidence from

our lab indicates that the neural representation of multiple speeds of overlapping stimuli in MT aligns with a global pooling scheme (Huang et al., 2023). Understanding these variations could provide deeper insights into the neural representation of multiple visual stimuli.

Modulation of MT response by object-based (surface-based) attention

Behavioral effects of object-based attention have been demonstrated using transparent motion stimuli (Valdes-Sosa et al., 1998, 2000; Mitchell et al., 2003, 2004). Using similar stimuli, Wannig et al. (2007) showed that the responses of MT neurons can be modulated by object-based attention directed to a moving surface of two overlapping stimuli. Modulation of neuronal responses by object-based attention has also been found in V1 (Roelfsema et al., 1998) and V4 (Fallah et al., 2007). Our study showed new evidence of response modulation by object-based attention in MT. Importantly, our visual stimuli and experimental paradigm allowed us to distinguish the effect of object-based attention from feature-based attention.

Responses of MT neurons can be modulated by feature-based attention to motion direction (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004), and binocular disparity (Ruff and Born, 2015). At the neuron population level, attending to the preferred direction of an MT neuron enhances the response to visual stimuli in the RF relative to attending to the null direction, regardless of whether the RF stimulus moves in the preferred or null direction of the neuron (Martinez-Trujillo and Treue, 2004). The same result was found for attending to binocular disparity (Ruff and Born, 2015). These results are consistent with the "feature-similarity gain (FSG) model" of attention (Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006). In our study, however, we found that attending to the more preferred disparity of MT neurons did not always enhance MT responses to two moving surfaces. When the surface at a more preferred disparity moved in a less preferred direction and elicited a weaker component response, attending to the more preferred disparity reduced MT response. Conversely, when the surface at a less preferred disparity moved in the preferred direction and elicited a stronger component response, attending to the less preferred disparity enhances MT response (Fig. 10). Therefore, the attention modulation was not due to feature-based attention to binocular disparity. To account for the neural response modulation found in our results, the attention initially directed to a cued binocular disparity of a stationary surface must spread to the motion direction of the attended surface of the overlapping stimuli, although the motion direction was never explicitly cued. Because the motion direction and the cued binocular disparity belong to the same surface, selecting the surface by attention would allow attention to spread to other features of the same object. Whether attention enhanced or reduced the neural response to two surfaces was dependent on whether the attended surface elicited a stronger or weaker component response than the other surface, rather than whether the attended surface was at the preferred or less preferred disparity of the neuron.

Since stereoscopic depth can be considered a spatial dimension in 3D space, may the effect of attention modulation found in our study be due to spatial attention? In response to two 2D stimuli placed at different locations within the RF, the neuronal response in MT (Treue and Maunsell, 1996; Seidemann and Newsome, 1999), V2, and V4 (Moran and Desimone, 1985; Luck et al., 1997; Reynolds et al., 1999) is enhanced (or reduced) when attention is directed to the spatial location where a stimulus component is a preferred (or null) stimulus of the neuron when presented alone. It has been shown that these results can be explained by the biased competition model (Desimone and Duncan, 1995; Reynolds et al., 1999) and the model of divisive normalization

(Ghose and Maunsell, 2008; Lee and Maunsell, 2009; Reynolds and Heeger, 2009; Ni et al., 2012). Our finding that attending to a surface eliciting a stronger (or weaker) component response when presented alone enhanced (or reduced) the response to overlapping 3D stimuli is consistent with these previous studies. Our study made two new advancements. First, it extended the finding to 3D stimuli with attention directed to the feature of binocular disparity. Second, previous studies using 2D stimuli did not measure the spatial preferences of the neurons. In contrast, our study directed animals' attention to the more preferred or less preferred disparity of the neurons and showed that the attention modulation was not dependent on the attended disparity, but dependent on the combined stimulus features of the attended surface as a whole. Regarding whether the attention modulation found in our study was due to spatial attention or object-based attention, we think the distinction may be more semantic than mechanistic. One can think of a surface in our study either as an object or a spatial location in depth. In the previous studies, although attention was directed to a spatial location within the RF, it was also possible that attention selected the visual stimulus as an object at the attended location which allowed spatial attention to spread to other features of the stimulus such as motion direction or orientation. Cavanagh and colleagues (2023) have recently suggested that spatial attention may rely on object-based attention.

Functional implication of disparity bias on segmentation and figure-ground segregation

It becomes easier to segment two overlapping surfaces when they are moving at different depths than at the same depth (Hibbard and Bradshaw, 1999; Greenwood and Edwards, 2006). The neural representation of two surfaces moving transparently in different directions and at different depths is distributed in the responses of populations of neurons that are selective for motion directions and binocular disparities. Without disparity cues, the visual system can only rely on motion cues to segment the overlapping surfaces (Xiao and Huang, 2015). When two moving surfaces are located at different depths, even when neurons simply average the component responses without showing disparity bias, MT neurons that prefer the disparity of one surface would still carry more information about the motion direction of that surface and therefore aid in segmentation. However, the benefit can be limited when the differences in stimulus features (disparity and direction) are small. The disparity bias that we found would further enhance the neural representation of the motion direction at the surface located at the more preferred disparity of the neurons. This disparity bias, independent of attention modulation, could help further partition the neural representation of multiple stimuli into distributed responses across different neuronal populations and therefore facilitate segmentation. Together, these mechanisms may provide a neural basis for the perceptual benefit of depth cues on segmentation.

Three lines of evidence from our results suggest that the near surface of overlapping stimuli may be better represented than the far surface. First, in monkey G, we found a predominant near bias across near and far neurons. The overall near bias could facilitate the segmentation of the near surface, which is consistent with the better performance of monkey G in discriminating motion directions of the near surface than the far surface of overlapping stimuli. Second, for animals like monkeys B and R, near surface may be better represented than the far surface because more neurons tend to prefer near disparity over far disparity across visual areas of V1, V2, V4 (Prince et al., 2002; Chen et al. 2008; Gonzalez et al., 2010; Samonds et al., 2012; Hinkle et al., 2005; Tanabe et al., 2005), and MT and MST (Maunsell and Van Essen, 1983b; DeAngelis and Uka, 2003; Gonzalez et al., 2001). Since near neurons tend to show near bias and far neurons tend to show far bias, having more near neurons in the population would better represent the near surface

of overlapping stimuli. Third, when combining all neurons in our dataset for each of the three animals, we found an initial near bias relative to the average of the component responses (Fig. 9, Table 1). This finding suggests that, at least in the early response, the near surface is better represented than the far surface, making it likely easier for the near surface to be segregated from the far surface, or to be selected by attention. Figural objects in natural scenes are more likely to reside at near disparity than far disparity relative to the ground region (Huang et al., 2019 SFN abstract). A better neural representation of the near surface in the visual cortex during the early neural response may reflect a strategy that the visual system uses to exploit the statistical regularity of natural vision and to better serve the function of figure-ground segregation. The current study only used surfaces at two disparities of -0.1° and 0.1°. Future studies using a range of disparities are needed to further test this idea.

Methods

Subjects and neural recording

Three adult male rhesus monkeys (Macaca mulatta) participated in the behavioral and neurophysiological experiments. All experimental procedures were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Wisconsin–Madison, adhering to U.S. Department of Agriculture regulations and National Institutes of Health guidelines for the care and use of laboratory animals. The procedures for surgical preparation and electrophysiological recordings were similar to those described in previous studies (Xiao and Huang, 2015).

We implanted a head post and a recording cylinder during sterile surgery, with the animal under isoflurane anesthesia. For electrophysiological recordings from neurons in area MT, we employed a vertical approach, using either single tungsten electrodes (1–3 MΩ, FHC) or 16-channel linear probes (AlphaOmega; Plexon). We identified area MT in each animal using structural MRI, along with its hallmark features: a high proportion of direction-selective neurons, smaller RFs compared to those in the adjacent medial superior temporal cortex (area MST), its specific anatomical location on the posterior bank of the superior temporal sulcus, and the visual topography of the RFs (Gattass and Gross, 1981). We amplified the electrical signals and identified single units using a real-time template-matching system in conjunction with an offline spike sorter (Plexon). Eye position was monitored at a rate of 1000 Hz using a video-based eye tracker (EyeLink, SR Research). In certain recording sessions with monkey B, we tracked both the left and right eye positions, which were subsequently used to calculate horizontal vergence by subtracting the horizontal position of the right eye from that of the left eye.

Experimental procedures

We used the real-time data acquisition program "Maestro" to control both stimulus presentation and data collection (https://sites.google.com/a/srscicomp.com/maestro). Visual stimuli were displayed on a 25-inch CRT monitor (SONY) positioned at a viewing distance of 63 cm. The monitor had a resolution of 1024 × 768 pixels and a refresh rate of 100 Hz. Stimuli were generated on a Linux workstation using an OpenGL application that communicated with the experimental control computer. Luminance calibration of the CRT monitor was performed using

a photometer (Minolta LS-110). Gamma correction was applied and the output linearity was subsequently confirmed.

Once the spike waveforms of a single neuron were isolated, we initially assessed the neuron's direction selectivity by presenting interleaved trials of a 30° × 27° random-dot patch (RDP), which moved in various directions with 45° increments at a speed of 10°/s. The direction selectivity and preferred direction (PD) were characterized in real time using MATLAB (MathWorks) for data analysis and visualization. Next, we assessed the neuron's speed tuning by presenting a random dot patch moving at various speeds (1, 2, 4, 8, 16, 32, or 64°/s) in the neuron's PD. We determined the preferred speed (PS) by fitting the speed-tuning curve with a cubic spline and identifying the speed that elicited the highest firing rate. Following this, we mapped the neuron's RF by recording responses to a series of 5° × 5° random dot patches moving in the PD at the PS. These patches were systematically positioned across the screen (35° × 25°) in 5° steps without overlap. We then interpolated the raw response map at 0.5° intervals to identify the location that produced the highest firing rate, designating this point as the center of the RF.

We used green and red anaglyph filters (Kodak Wratten filters, #25 and 58) to generate stereoscopic depth on the monitor through horizontal disparity. To create a disparity of +n° (far) or -n° (near), we presented two identical dot patterns, one in green and the other in red, with the center of the red dot pattern offset horizontally by +n° or -n° relative to the green dot pattern. When the animals viewed the stimuli through a green filter over the left eye and a red filter over the right eye, each eye perceived one set of dots, producing a binocular disparity that resulted in depth perception. Crosstalk between the two eyes, as viewed through the filters, was measured at 14.5%, 9.6%, and 9.6% for monkeys G, B, and R, respectively. Crosstalk was calculated as the luminance of red (green) dots viewed through the green (red) filter divided by the luminance of red (green) dots viewed through the red (green) filter, with the listed values representing the average crosstalk for red and green dots. The crosstalk remained relatively high even when using two layers of Kodak Wratten filters. Upon investigating the source of the crosstalk, we found that the CRT monitor emitted a faint white light even when the look-up table (LUT) value for background luminance was set to zero. The background luminance measured through the green and red filters were 1.01 and 0.68 cd/m², respectively. This suggests that the light leaking through the anaglyph filter of one color was likely ambient background light, rather than the dots of the opposite color. Human observers, using the same anaglyph filters, confirmed the perception of clear stereoscopic depth with two overlapping surfaces at different binocular disparities. We evaluated an MT neuron's selectivity to binocular disparity by recording its responses to a series of random dot patterns moving in the neuron's PD and at its PD speed (typically between 5° and 20°/s). These RDPs were presented at fifteen different disparities, ranging from -1.5° to +1.5°.

Visual stimuli

Our primary visual stimuli consisted of two overlapping RDPs moving at different depths within a stationary circular aperture with a 15° diameter. One RDP moved at a binocular disparity of -0.1° (near depth) and the other at 0.1° (far depth) (Fig. 2A-D). The direction separation between the two moving surfaces was set at either 60° or 120°. We refer to each RDP as a "stimulus component," with one being the near component and the other the far component. The dot density for each RDP was 0.89 dots/deg², and both RDPs moved at 100% coherence. As each dot reached the border of the aperture, it would disappear and reappear on the opposite side, continuing its motion in the same direction until the end of the motion period. We varied the VA direction of the

bi-directional stimuli in randomly interleaved trials. We also randomly interleaved trials in which only the near or far component was shown alone.

To maximize the behavioral performance of monkeys B and G on the 12AFC task, we typically centered the visual stimuli on the fixation point. Sometimes, we shifted the stimuli 1 to 2° towards the RFs to better cover the RFs. As a result, the visual stimuli encompassed the MT RFs, though they were not directly centered on them. For monkey R, the bi-directional stimuli and component stimuli were centered on the RFs with a diameter of 15°. Meanwhile, a task-relevant random-dot stimulus with a diameter of 5°, moving in a single direction at the fixation plane, was presented in the opposite visual hemifield, centered 10° to the left of the fixation spot (see behavioral tasks below). The speed of the motion stimuli was typically set at 10°/s, with occasional adjustments to 12°/s. These stimulus speeds drove most MT neurons effectively and facilitated visual segmentation of transparently moving stimuli at these relatively slow speeds.

Behavioral tasks

Monkeys B and G performed a 12-alternative forced-choice (12AFC) direction discrimination task to report the motion direction of a surface at a cued depth. Each trial began with the animals fixating on a central point, while stationary reference dots at zero disparity were displayed outside the region where the stimuli would later appear. The monkeys had to maintain fixation within a 1° × 1° window. To direct attention, a static circular random dot pattern (RDP) with a 15° diameter appeared for 500 ms at either near (-0.1°) or far (0.1°) disparity. Stationary reference dots at zero disparity filled the rest of the display to enhance depth perception. Following a 300-ms blank screen with the fixation point, a single RDP at the cued disparity or two overlapping RDPs at 0.1° and -0.1° disparities appeared stationary for 500 ms, followed by 600 ms of motion. The VA direction of the motion was one of 12 directions spaced at 30° intervals from 0° to 360°. After the motion period, the stimuli disappeared, and 12 reporting targets, arranged in a 360° circle at 30° intervals with a 10° diameter, appeared for 450 ms for Monkey B and 600 ms for Monkey G. During this time, the monkeys reported the motion direction of the cued component by making a saccade to the corresponding target in the matching direction. They were required to hold fixation within a 1.5° × 1.5° window around the correct target for 350 ms to receive a juice reward.

In the task where attention was directed away from the RFs, single- and bi-directional stimuli at different depths were centered on the RFs of MT neurons, while Monkey R attended to a task-relevant stimulus in the opposite visual hemifield. Monkey R performed a 2AFC fine direction discrimination task, reporting whether the motion of an RDP was clockwise or counter-clockwise relative to an invisible upward vertical. The monkey maintained fixation within a 1° × 1° window until making a saccade to report. The attended stimulus was a 5° diameter RDP, placed 10° to the left of fixation at zero disparity, while RF stimuli were positioned in the right hemifield (recording from the left MT). Each trial began with a fixation point and reference stationary dots at zero disparity. After 200 ms of fixation, both the RF and attended stimuli appeared stationary for 250 ms before moving for 500 ms. The attended stimulus moved at 5°/s, either clockwise or counter-clockwise from vertical by offsets of 10°, 15°, or 20°. The RF stimuli with either a single RDP or overlapping RDPs moved in one of the 12 directions as described above in "Visual stimuli". All trials were randomly interleaved. After the motion period, stimuli were turned off, and two reporting targets appeared 10° to the left and right of fixation. Monkey R had 400 ms to saccade

to the left target if the attended motion was counter-clockwise to the upward vertical, or the right target if clockwise to vertical, to receive a juice reward.

Data analysis

Neuronal responses were quantified as the firing rate (spikes per second) during the stimulus motion period, averaged across repeated trials. We constructed tuning curves for both unidirectional and bi-directional stimuli, fitting the raw data using cubic splines with 1° interpolation steps. For each VA direction of the overlapping stimuli, we measured the responses to the bi-directional stimuli as well as to the individual unidirectional components. To average the direction tuning curves across neurons, we aligned the spline-fitted tuning curves for bi-directional stimuli by rotating them so that the VA direction of 0° corresponded to each neuron's PD. We then normalized the responses based on the maximum firing rate to either the near or far component, whichever was greater, and averaged the aligned, normalized tuning curves across all neurons.

We calculated the population-averaged tuning curves by combining responses from the Near C/Far CC and Near CC/Far C conditions (Fig. 3). For visualization, the near component was positioned on the clockwise side and the far component on the counter-clockwise side. To achieve this, we horizontally flipped the spline-fitted, aligned tuning curves for each neuron in the Near CC/Far C condition relative to the vertical axis at VA direction 0° (the neuron's PD). The flipped tuning curves were then averaged with the tuning curves from the Near C/Far CC condition.

To analyze the time course of response tuning to the bi-directional stimuli, we calculated the tuning curves for each neuron using trial-averaged firing rates within a sliding 50-ms window, stepping by 10 ms. Responses were normalized to the maximum firing rate across all time bins, based on the stronger response to either the near or far component. We then pooled the responses from the Near C/Far CC and Near CC/Far C conditions, as described above, and averaged the normalized responses across neurons.

To evaluate the goodness of fit for a model of response tuning to bi-directional stimuli, we calculated the percentage of variance (PV) explained by the model using the following formula:

$$PV = 100 \times (1 - \frac{SSE}{SST}) \qquad (13)$$

where SSE is the sum of squared errors between the model fit and the neural data, and SST is the sum of squared differences between the data and the mean of the data.

Divisive normalization model (DNM)

Probability density function of weighting pool size

We fitted each neuron's response tuning curve to the bi-directional stimuli using an extended divisive normalization model. Each neuron contributed to the fits of eight bi-directional tuning curves presented in the results (2 direction separations of 60° and 90° × 2 stimulus configurations of "Near C/Far CC" and "Near CC/Far C" × 2 attention conditions of "Attend Near" and "Attend Far"). Equations 11 and 12 were used to fit the bi-directional tuning curves of each neuron. When fitting the response tuning curve with the DNM model, we divided the parameter

search space for the parameter v (the standard deviation of the Gaussian function $\boldsymbol{G}$) into steps that resulted in equal steps of the area under the truncated Gaussian function $\boldsymbol{G}$ (Eq. 10) (within -1.5° to 1.5° disparities) (Eq. 10). We did so because the area under $\boldsymbol{G}$ better represents the weighting pool size than v, given the nonlinear relationship between v and the area. For each v value in the search space (ranging from 0.01 to 40 in 20 steps), we used MATLAB function "fmincon" to minimize the sum of squared errors (SSE) between the model fit and the neural data (Eqs. 11 and 12). The steps of v were equilvalent to even steps of normalized area under $\boldsymbol{G}$ from 0 to 1 in a step of 0.05.

We calculated the probability distribution of the weighting pool size that gave rise to the best fit of the data (measured by SSE) using the DNM. To do so, we first calculated the probability distribution of v, then converted it to the probability distribution of the area under the Gaussian function $\boldsymbol{G}$. For each neuron, we first calculated the SSE of the model fit as a function of v. Instead of selecting a single v value with the lowest SSE, we considered all values of v that produced SSE within the lowest 5%, as multiple v values could yield similar fits. Given the inherent noise in neuronal responses, it was important to account for all v values that provided a good fit. For each tuning curve, we normalized SSE (v) by the maximum SSE and defined a function $f(v)$ such that $f(v)$ = 1 when the normalized SSE at those v values were ≤ 0.05, and $f(v)$ = 0 at other v values. We then normalized $f(v)$ by its total area to obtain $p(v) = \frac{f(v)}{\Sigma f(v)}$. The probability density function $pdf(v)$ was computed by averaging $p(v)$ across all tuning curves from each neuron and further averaged across near or far neurons. Finally, we converted $pdf(v)$ to the probability density function based on the area under the Gaussian function $\boldsymbol{G}$ (Fig. 11B and C).

To evaluate whether $pdf(v)$ differed between the monkeys, we used bootstrapping where we randomly sampled the $pdf(v)$ from a neuronal set of either the near neurons or far neurons 1000 times with replacement and calculated the 95% confidence intervals for each v value.

Estimate model parameters for the effect of attention

We derived the values of attention parameters β_{An} and β_{Af} used in Equations 11 and 12 using the following approach. We first fitted the response tuning to the bi-directional stimuli with the LWS model:

$$R_k(\theta_{na}, \theta_f) = w_{nAn} R_k(\theta_n) + w_{fAn} R_k(\theta_f) + c \quad , \qquad (14a)$$

$$R_k(\theta_n, \theta_{fa}) = w_{nAf} R_k(\theta_n) + w_{fAf} R_k(\theta_f) + c \quad , \qquad (14b)$$

and empirically determined the parameters w_{nAn}, w_{fAn}, w_{nAf}, and w_{fAf}, which are the weights for the near and far components in the two attention conditions. The first subscripts n and f refer to the near and far components, respectively. The subsequent subscripts An and Af denote attending near and attending far components, respectively. The parameter c is positive.

We assumed that attention to a stimulus component of the bi-directional stimuli multiplicatively scaled the response weight of the attended component (i.e. scaled by factor β_{An} or β_{Af}), but did not change the weight of the unattended component (i.e. scaled by a factor of 1).

So w_{nAn}, w_{fAn}, w_{nAf}, and w_{fAf} can be expressed as the following. When attention was directed to the near component:

$$w_{nAn} = \beta_{An} \cdot w_n \,, \qquad (15a)$$

$$w_{fAn} = w_f \,. \qquad (15b)$$

When attention was directed to the far component:

$$w_{nAf} = w_n \,, \qquad (15c)$$

$$w_{fAf} = \beta_{Af} \cdot w_f \,. \qquad (15d)$$

w_n and w_f are the response weights for the near and far components, respectively, when attention is not involved. Therefore, by combining Equations 15a and 15c, we can determine β_{An} (Eq. 16a).

$$\beta_{An} = \frac{w_{nAn}}{w_{nAf}} \,. \qquad (16a)$$

The attention factor β_{An} is essentially a ratio between the response weights for the near component under the attending near and attending far conditions. Since w_{nAn} and w_{nAf} can be estimated from the data, β_{An} can be determined accordingly. Similarly, by combining Equations 15d and 15b, we can determine β_{Af} (Eq. 16b).

$$\beta_{Af} = \frac{w_{fAf}}{w_{fAn}} \,. \qquad (16b)$$

Supplementary Materials

Methods and Results

We performed additional analysis to examine the timecourse of disparity bias and its relationship to neuronal responses. To track the neuronal response to bi-directional stimuli (R_{12}), we averaged the responses from the 'Attend Near' and 'Attend Far' conditions for monkeys B and G, and used the 'Attend Away' response for monkey R. The firing rate was calculated using a 50-ms sliding window, advancing in 10-ms steps. Within each time window centered on time point *t*, we integrated the neuronal response within a 30° direction window centered on the VA direction at the "Near PD" and "Far PD" regions of the neural responses from monkeys B, G, and R, as shown in the colormaps in Figures 6, 7, 8. The integrated responses, denoted as *R(NearPD, t)* and *R(FarPD, t)*, were averaged to estimate the timecourse of the overall neuronal response (blue curves in Supplementary Figures 1-3).

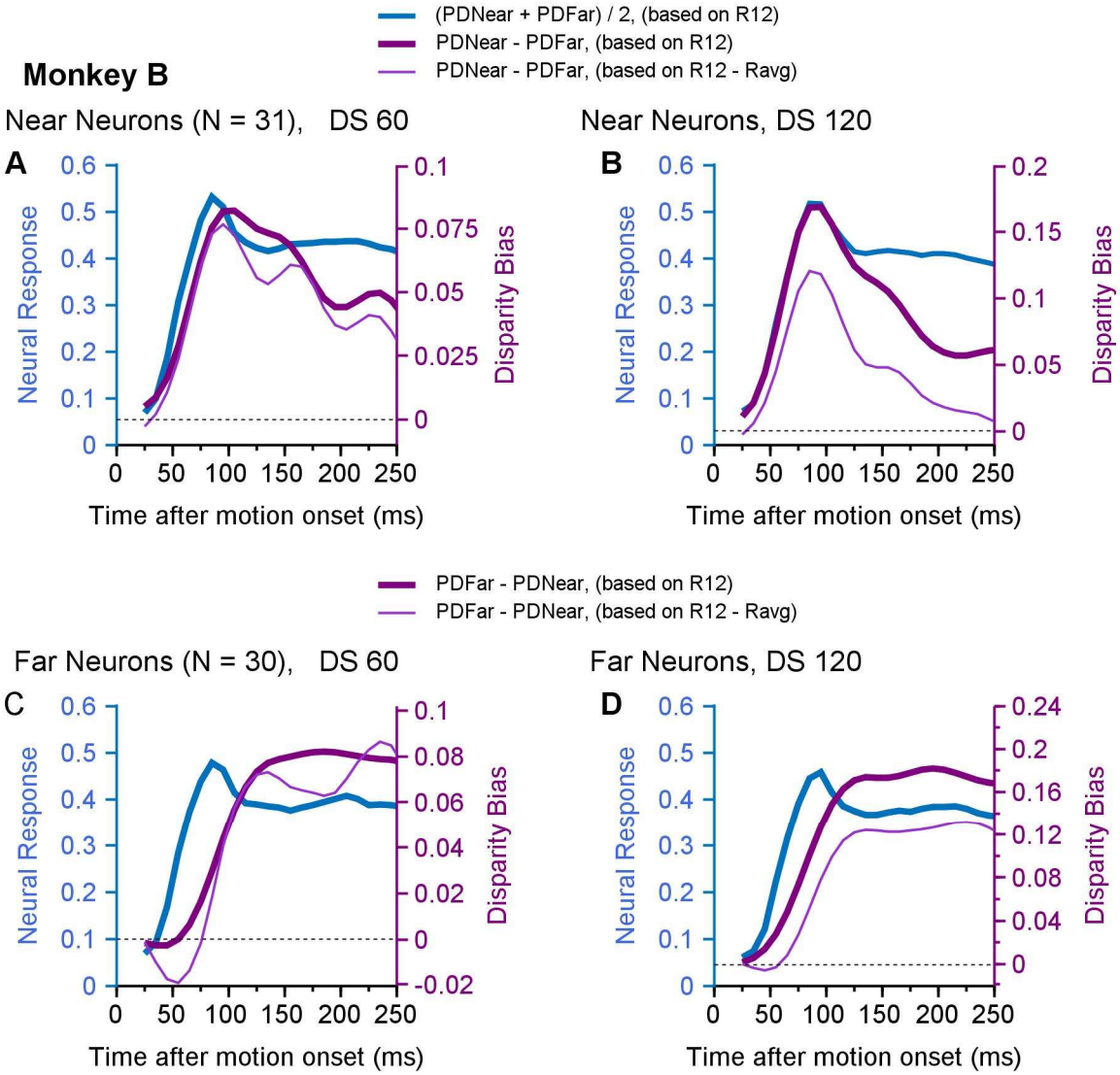
To capture the timecourse of disparity bias, we calculated the difference between *R(NearPD, t)* and *R(FarPD, t)* (purple curves in Supplementary Figures 1, 2, and 3). *R* was derived either from the bi-directional response R_{12} (thick purple lines) or from the difference between R_{12} and the average of the component responses R_{avg}, (i.e. R_{12} - R_{avg}) (thin purple lines).

For monkey B, relative to the neuronal response, the near bias of the near neurons (SFig. 1A, B) appeared sooner than the far bias of the far neurons (SFig. 1C, D). The disparity bias also appeared sooner at 120° DS (DS) (SFig. 1B, D) than at 60° DS (SFig. 1A, C). At 120° DS, the near bias of the near neurons almost had no delay (SFig. 1B). Interestingly, at 60° DS, the far neurons initially showed a weak near bias before changing to the far bias (the thin purple curve in SFig. 1C).

For monkey G, the neuronal response was more delayed relative to monkey B. The disparity bias of monkey G was also substantially more delayed (SFig. 2) than monkey B (SFig. 1). Both near neurons and far neurons of monkey G showed near bias. Similar to monkey B, the near neurons of monkey G (SFig. 2A, B) showed a sooner disparity bias than the far neurons (SFig. 2C, D).

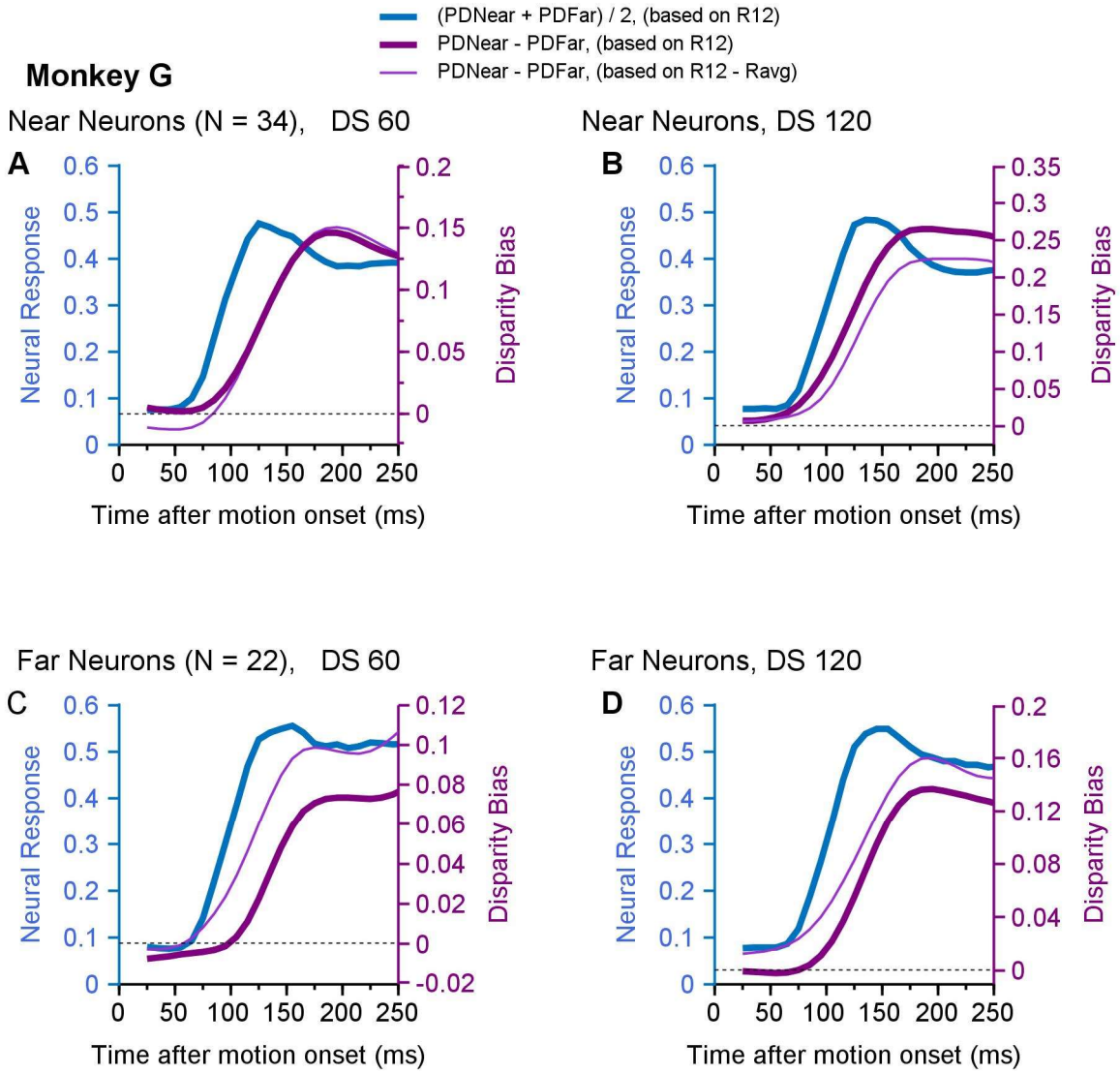
Similarly, the near neurons of monkey R showed a sooner near bias (SFig. 3A, B) than the far neurons showing a far bias (SFig. 3C, D). The near bias of the near neurons also rose faster than the far bias of the far neurons (comparing SFig. 3A with 3C, and 3B with 3D). As found in monkey B, the far neurons from monkey R initially showed a near bias, before gradually changing to the far bias (thin purple curves in SFig. 3C, 3D).

Supplementary Figure 1



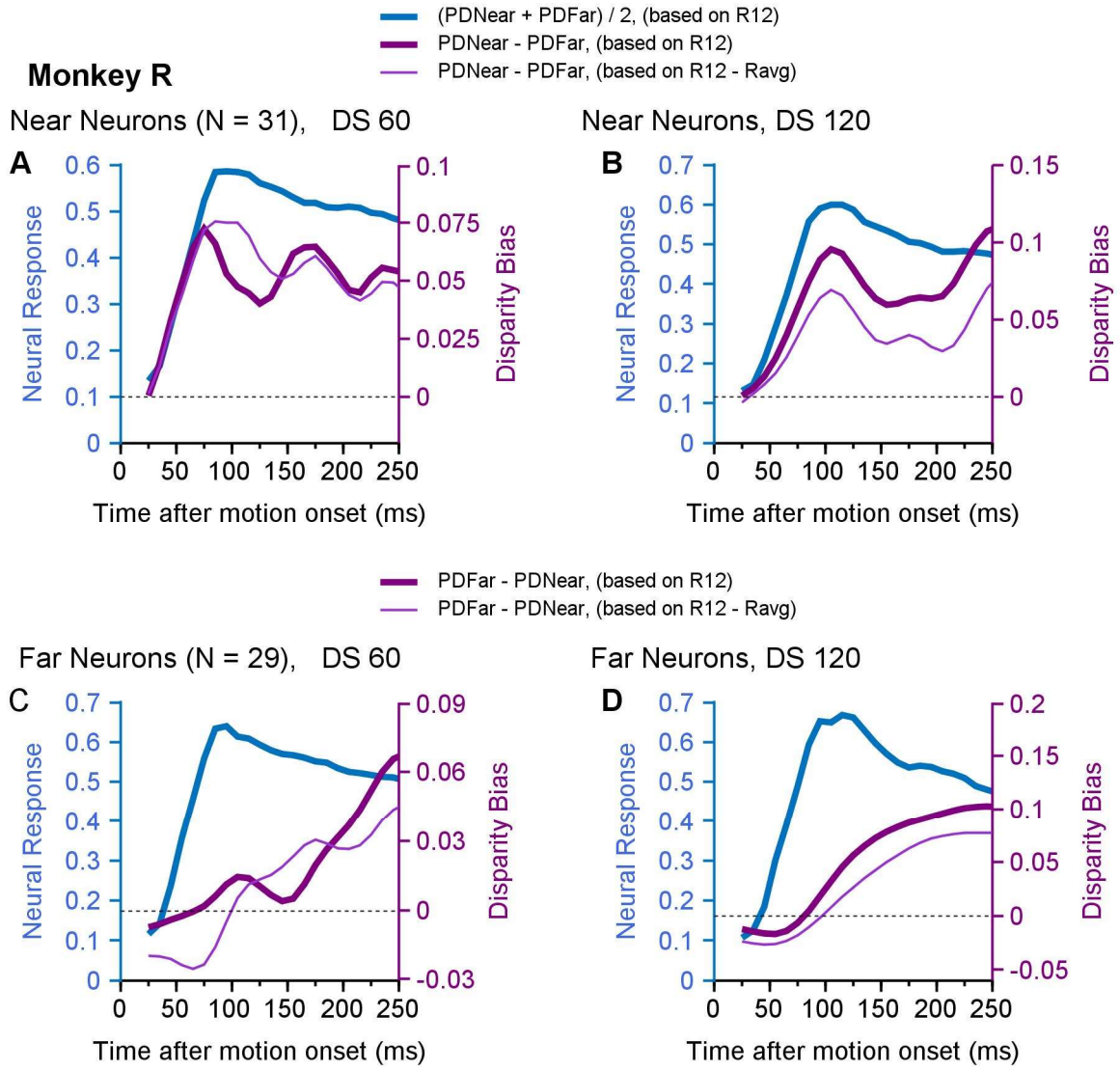
SFigure 1. Comparison of the timecourses of neuronal responses and disparity bias in monkey B. The blue curves show the population averaged neuronal response to bi-directional stimuli (*R12*) measured by averaging *R(NearPD)* and *R(FarPD).* The thick puple curves indicate the disparity bias by comparing *R(NearPD)* and *R(FarPD).* The thin puple curves indicate the disparity bias based on *R12 – Ravg*. The magnitude of the disparity bias is shown by the Y-axis on the right side of each plot, marked by purple color. **A, B**. Responses from near neurons. Disparity bias was calculated by *R(NearPD) - R(FarPD).* **C, D**. Responses from far neurons. Disparity bias was calculated by *R(FarPD) - R(NearPD).* Negative disparity bias below the dotted lines in C, D, indicates near bias. **A, C**. 60° direction separation. **B, D**. 120° direction separation.

Supplementary Figure 2



SFigure 2. Comparison of the timecourses of neuronal responses and disparity bias in monkey G. The convention is the same as Sfigure 1. **A, B**. Responses from near neurons. **C, D**. Responses from far neurons. Because the far neurons of monkey G also showed near bias, the disparity bias in C, D was calculated by *R(NearPD) - R(FarPD).* **A, C**. 60° direction separation. **B, D**. 120° direction separation.

Supplementary Figure 3



SFigure 3. Comparison of the timecourses of neuronal responses and disparity bias in monkey R. The convention is the same as Sfigure 1. **A, B**. Responses from near neurons. Disparity bias was calculated by *R(NearPD) - R(FarPD).* **C, D**. Responses from far neurons. Disparity bias was calculated by *R(FarPD) - R(NearPD).* Negative disparity bias below the dotted lines in C, D, indicates near bias. **A, C**. 60° direction separation. **B, D**. 120° direction separation.

References

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), 1106–1130. https://doi.org/10.1152/jn.1984.52.6.1106

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), 1774. https://doi.org/10.1038/s41467-018-04126-7

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*(1), 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Bradley, D. C., Chang, G. C., & Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, *392*(6677), 714–717. https://doi.org/10.1038/33688

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, *373*(6515), 609–611. https://doi.org/10.1038/373609a0

Britten, K. H. (2003). The Middle Temporal Area: Motion Processing and the Link to Perception. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences, 2-vol. Set* (pp. 1203–1216). The MIT Press. https://doi.org/10.7551/mitpress/7131.003.0093

Britten, K. H., & Heuer, H. W. (1999). Spatial summation in the receptive fields of MT neurons. *The Journal of Neuroscience*, *19*(12), 5074–5084. https://doi.org/10.1523/JNEUROSCI.19-12-05074.1999

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*(12), 4745–4765. https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*(6), 1157–1169. https://doi.org/10.1017/s0952523800010269

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, *64*(6), 931–942. https://doi.org/10.1016/j.neuron.2009.11.004

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51–62. https://doi.org/10.1038/nrn3136

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*, *17*(21), 8621–8644. https://doi.org/10.1523/JNEUROSCI.17-21-08621.1997

Cavanagh, P., Caplovitz, G. P., Lytchenko, T. K., Maechler, M. R., Tse, P. U., & Sheinberg, D. L. (2023). The Architecture of Object-Based Attention. *Psychonomic Bulletin & Review*, *30*(5), 1643–1667. https://doi.org/10.3758/s13423-023-02281-7

Chen, G., Lu, H. D., & Roe, A. W. (2008). A map for horizontal disparity in monkey V2. *Neuron*, *58*(3), 442–450. https://doi.org/10.1016/j.neuron.2008.02.032

Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology*, *71*(6), 2269–2280. https://doi.org/10.1152/jn.1994.71.6.2269

Cumming, B. G., & DeAngelis, G. C. (2001). The Physiology of Stereopsis. *Annual Review of Neuroscience*, *24*(Volume 24, 2001), 203–238. https://doi.org/10.1146/annurev.neuro.24.1.203

Cumming, B. G., & Parker, A. J. (1999). Binocular Neurons in V1 of Awake Monkeys Are Selective for Absolute, Not Relative, Disparity. *Journal of Neuroscience*, *19*(13), 5602–5618. https://doi.org/10.1523/JNEUROSCI.19-13-05602.1999

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of

stereoscopic depth. *Nature*, *394*(6694), 677–680. https://doi.org/10.1038/29299

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *The Journal of Neuroscience*, *19*(4), 1398–1415. https://doi.org/10.1523/JNEUROSCI.19-04-01398.1999

DeAngelis, G. C., & Uka, T. (2003). Coding of Horizontal Disparity and Velocity by MT Neurons in the Alert Macaque. *Journal of Neurophysiology*, *89*(2), 1094–1111. https://doi.org/10.1152/jn.00717.2002

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually Bistable Three-Dimensional Figures Evoke High Choice Probabilities in Cortical Area MT. *Journal of Neuroscience*, *21*(13), 4809–4821. https://doi.org/10.1523/JNEUROSCI.21-13-04809.2001

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, *113*(4), 501–517. https://doi.org/10.1037//0096-3445.113.4.501

Fallah, M., Stoner, G. R., & Reynolds, J. H. (2007). Stimulus-specific competitive selection in macaque extrastriate visual area V4. *PNAS*, *104*(10), 4165–4169. https://doi.org/10.1073/pnas.0611722104

Ferrera, V. P., & Lisberger, S. G. (1997). Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *Journal of Neurophysiology*, *78*(3), 1433–1446. https://doi.org/10.1152/jn.1997.78.3.1433

Ghose, G. M., & Maunsell, J. H. R. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *The Journal of Neuroscience*, *28*(19), 5115–5126. https://doi.org/10.1523/JNEUROSCI.0138-08.2008

Gibson, J. J. (1950). *The perception of the visual world*. Houghton Mifflin.

Gonzalez, F., Bermudez, M. A., Vicente, A. F., & Romero, M. C. (2010). Orientation preference and horizontal disparity sensitivity in the monkey visual cortex. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *30*(6), 824–833. https://doi.org/10.1111/j.1475-1313.2010.00781.x

Gonzalez, F., Perez, R., Justo, M. S., & Bermudez, M. A. (2001). Receptive field organization of disparity-sensitive cells in Macaque medial superior temporal cortex. *The European Journal of Neuroscience*, *14*(1), 167–173. https://doi.org/10.1046/j.0953-816x.2001.01626.x

Goutcher, R. (2016). Motion direction influences surface segmentation in stereo transparency. *Journal of Vision*, *16*(15), 17. https://doi.org/10.1167/16.15.17

Greenwood, J. A., & Edwards, M. (2006). Pushing the limits of transparent-motion detection with binocular disparity. *Vision Research*, *46*(16), 2615–2624. https://doi.org/10.1016/j.visres.2006.01.022

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181–197. https://doi.org/10.1017/s0952523800009640

Heuer, H. W., & Britten, K. H. (2002). Contrast dependence of response normalization in area MT of the rhesus macaque. *Journal of Neurophysiology*, *88*(6), 3398–3408. https://doi.org/10.1152/jn.00255.2002

Hibbard, P. B., & Bradshaw, M. F. (1999). Does Binocular Disparity Facilitate the Detection of Transparent Motion? *Perception*, *28*(2), 183–191. https://doi.org/10.1068/p2742

Hinkle, D. A., & Connor, C. E. (2005). Quantitative characterization of disparity tuning in ventral pathway area V4. *Journal of Neurophysiology*, *94*(4), 2726–2737. https://doi.org/10.1152/jn.00341.2005

Huang, X., Ghimire, B., Chakrala, A. S., & Wiesner, S. (2023). Neural encoding of multiple motion speeds in visual cortical area MT. *bioRxiv: The Preprint Server for Biology*, 2023.04.08.532456. https://doi.org/10.1101/2023.04.08.532456

Huang, X., Wang, C., Arseneau, B., Yerxa, T.E., Cooper, E.A. (2019). Natural scene statistics of depth and motion pertinent to figure-ground segregation. *Society for Neuroscience Abstracts*, 487, 7048.

Kouh, M., & Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Computation*, *20*(6), 1427–1451. https://doi.org/10.1162/neco.2008.02-07-466

Krug, K., Cicmil, N., Parker, A. J., & Cumming, B. G. (2013). A Causal Role for V5/MT Neurons Coding Motion-Disparity Conjunctions in Resolving Perceptual Ambiguity. *Current Biology*, *23*(15), 1454–1459. https://doi.org/10.1016/j.cub.2013.06.023

Krug, K., & Parker, A. J. (2011). Neurons in Dorsal Visual Area V5/MT Signal Relative Disparity. *Journal of Neuroscience*, *31*(49), 17892–17904. https://doi.org/10.1523/JNEUROSCI.2658-11.2011

Lee, J., & Maunsell, J. H. R. (2009). A normalization model of attentional modulation of single unit responses. *PloS One*, *4*(2), e4651. https://doi.org/10.1371/journal.pone.0004651

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*(1), 24–42. https://doi.org/10.1152/jn.1997.77.1.24

Manning T. S., Alexander E., Cumming C. G., DeAngelis G. C., Huang X., and Cooper E. A. (2024). Transformations of sensory information in the brain reflect a changing definition of optimality. *PLOS Computational Biology.* https://doi.org/10.1371/journal.pcbi.1011783

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology: CB*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, *49*(5), 1127–1147. https://doi.org/10.1152/jn.1983.49.5.1127

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*(5), 1148–1167. https://doi.org/10.1152/jn.1983.49.5.1148

Mitchell, J. F., Stoner, G. R., Fallah, M., & Reynolds, J. H. (2003). Attentional selection of superimposed surfaces cannot be explained by modulation of the gain of color channels. *Vision Research*, *43*(12), 1323–1328. https://doi.org/10.1016/s0042-6989(03)00123-8

Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature*, *429*(6990), 410–413. https://doi.org/10.1038/nature02584

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Ni, A. M., & Maunsell, J. H. R. (2019). Neuronal Effects of Spatial and Feature Attention Differ Due to Normalization. *The Journal of Neuroscience*, *39*(28), 5493–5505. https://doi.org/10.1523/JNEUROSCI.2106-18.2019

Ni, A. M., Ray, S., & Maunsell, J. H. R. (2012). Tuned normalization explains the size of attention modulations. *Neuron*, *73*(4), 803–813. https://doi.org/10.1016/j.neuron.2012.01.006

Pasternak, T., & Tadin, D. (2020). Linking Neuronal Direction Selectivity to Perceptual Decisions About Visual Motion. *Annual Review of Vision Science*, *6*(1), 335–362. https://doi.org/10.1146/annurev-vision-121219-081816

Prince, S. J. D., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology*, *87*(1), 209–221. https://doi.org/10.1152/jn.00466.2000

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *Journal of Neuroscience*, *14*(12), 7357–7366. https://doi.org/10.1523/JNEUROSCI.14-12-07357.1994

Recanzone, G. H., & Wurtz, R. H. (2000). Effects of attention on MT and MST neuronal activity during pursuit initiation. *Journal of Neurophysiology*, *83*(2), 777–790. https://doi.org/10.1152/jn.2000.83.2.777

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience*, *19*(5), 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381. https://doi.org/10.1038/26475

Ruff, D. A., & Born, R. T. (2015). Feature attention for binocular disparity in primate area MT depends on tuning strength. *Journal of Neurophysiology*, *113*(5), 1545–1555. https://doi.org/10.1152/jn.00772.2014

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431. https://doi.org/10.1038/nn1786

Samonds, J. M., Potetz, B. R., & Lee, T. S. (2012). Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *PNAS*, *109*(16), 6313–6318. https://doi.org/10.1073/pnas.1200125109

Sclar, G., Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, *30*(1), 1–10. https://doi.org/10.1016/0042-6989(90)90123-3

Seidemann, E., & Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *Journal of Neurophysiology*, *81*(4), 1783–1794. https://doi.org/10.1152/jn.1999.81.4.1783

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*(5), 743–761. https://doi.org/10.1016/s0042-6989(97)00183-1

Snowden, R. J., & Rossiter, M. C. (1999). Stereoscopic Depth Cues Can Segment Motion Information. *Perception*, *28*(2), 193–201. https://doi.org/10.1068/p2735

Sprague W. S., Cooper E. A., Tošić I., Banks M. S. (2015). Stereopsis is adaptive for the natural environment. Science Adv.;1:e1400254. https://doi: 10.1126/sciadv.1400254

Tanabe, S., Doi, T., Umeda, K., & Fujita, I. (2005). Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *Journal of Neurophysiology*, *94*(4), 2683–2699. https://doi.org/10.1152/jn.00319.2005

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, *382*(6591), 539–541. https://doi.org/10.1038/382539a0

Uka, T., & DeAngelis, G. C. (2003). Contribution of Middle Temporal Area to Coarse Depth Discrimination: Comparison of Neuronal and Psychophysical Sensitivity. *Journal of Neuroscience*, *23*(8), 3515–3530. https://doi.org/10.1523/JNEUROSCI.23-08-03515.2003

Uka, T., & DeAngelis, G. C. (2004). Contribution of Area MT to Stereoscopic Depth Perception: Choice-Related Response Modulations Reflect Task Strategy. *Neuron*, *42*(2), 297–310. https://doi.org/10.1016/S0896-6273(04)00186-2

Uka, T., & DeAngelis, G. C. (2006). Linking Neural Representation to Function in Stereoscopic Depth Perception: Roles of the Middle Temporal Area in Coarse versus Fine Disparity Discrimination.

Journal of Neuroscience, *26*(25), 6791–6802. https://doi.org/10.1523/JNEUROSCI.5435-05.2006

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition*, *66*(2), B13–B23. https://doi.org/10.1016/s0010-0277(98)00012-2

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(2), 488–505. https://doi.org/10.1037//0096-1523.26.2.488

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wannig, A., Rodríguez, V., & Freiwald, W. A. (2007). Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron*, *54*(4), 639–651. https://doi.org/10.1016/j.neuron.2007.05.001

Wiesner, S., Baumgart, I. W., & Huang, X. (2020). Spatial Arrangement Drastically Changes the Neural Representation of Multiple Visual Stimuli That Compete in More Than One Feature Domain. *The Journal of Neuroscience*, *40*(9), 1834–1848. https://doi.org/10.1523/JNEUROSCI.1950-19.2020

Wiesner, S., Huang, X. (2019). Image segmentation and neural representation of multiple spatially-separated stimuli in cortical area MT. *Society for Neuroscience Abstracts.*

Xiao, J., & Huang, X. (2015). Distributed and Dynamic Neural Encoding of Multiple Motion Directions of Transparently Moving Stimuli in Cortical Area MT. *The Journal of Neuroscience*, *35*(49), 16180–16198. https://doi.org/10.1523/JNEUROSCI.2175-15.2015

Xiao, J., Niu, Y.-Q., Wiesner, S., & Huang, X. (2014). Normalization of neuronal responses in cortical area MT across signal strengths and motion directions. *Journal of Neurophysiology*, *112*(6), 1291–1306. https://doi.org/10.1152/jn.00700.2013