Psychophysical studies and our own subjective experience suggest that, in natural viewing conditions (i.e., at medium to high contrasts), monocularly and binocularly viewed scenes appear very similar, with the exception of the improved depth perception provided by stereopsis. This phenomenon is usually described as a lack of binocular summation. We show here that there is an exception to this rule: Ocular following eye movements induced by the sudden motion of a large stimulus, which we recorded from three human subjects, are much larger when both eyes see the moving stimulus, than when only one eye does. We further discovered that this binocular advantage is a function of the interocular correlation between the two monocular images: It is maximal when they are identical, and reduced when the two eyes are presented with different images. This is possible only if the neurons that underlie ocular following are sensitive to binocular disparity.

**Introduction**

The primary benefit of having two forward-facing eyes is enhanced depth perception. In principle, combining signals from both eyes could improve other functions, such as form detection. However, a large body of psychophysical research has revealed that, whereas at very low contrasts (around perception threshold) binocular vision confers a significant advantage in contrast sensitivity and visual acuity, this vanishes quickly as contrast is increased to the suprathreshold levels typical of our daily experience (Campbell & Green, 1965; Rose, 1980; Arditi, Anderson, & Movshon, 1981; Legge & Rubin, 1981; Legge, 1984a; Legge, 1984b; Anderson & Movshon, 1989; Caganello, Arditi, & Halpern, 1993; Zlatkova, Anderson, & Ennis, 2001; Ding & Sperling, 2006; Meese, Georgeson, & Baker, 2006; Baker, Meese, & Georgeson, 2007; Hess, Hutchinson, Ledgeway, & Mansouri, 2007; Pineles, Velez, Yu, Demer, & Birch, 2014). Similarly, binocular vision improves neither smooth pursuit eye movements (González, Lillakas, Greenwald, Gallie, & Steinbach, 2014; Shanidze, Heinen, & Verghese, 2017) nor saccadic eye movements (Krauskopf, Cornsweet, & Riggs, 1960). In fact, thanks to our exquisite ability to exploit monocular depth clues, under monocular viewing we even maintain a vivid sensation of depth. Accordingly, most human activities, including operating a motor vehicle, can be carried out by monocularly blind subjects.

Neural processing of binocular stimuli starts in primary visual cortex (area V1 in primates), where most neurons receive inputs from both eyes, and are more strongly activated by their preferred binocular stimulus than by their preferred monocular stimulus (Barlow, Blakemore, & Pettigrew, 1967; Pettigrew, Nikara, & Bishop, 1968; Poggio & Fischer, 1977; Prince, Pointon, Cumming, & Parker, 2002), at all contrast levels. Accordingly, perceptual invariance between monocular and binocular presentations requires either selective pooling of V1 responses, or additional processing beyond V1. Other behavioral responses might not share
such pooling/processing, and thus might reveal a clear binocular advantage. Anecdotal reports (Miles, Kawano, & Optican, 1986; Inoue, Takemura, Kawano, & Mustari, 2000) indicate that in monkeys reflexive eye movements known as ocular following responses (OFRs), supported primarily by the dorsal visual pathway (Miles, 1997; Miles, 1998; Kawano, 1999; Masson, 2004; Miles, Busettini, Masson, & Yang, 2004; Takemura, Murata, Kawano, & Miles, 2007), are stronger for binocular than monocular high contrast stimuli, suggesting that OFRs might exhibit binocular summation.

To test this hypothesis, we recorded OFRs induced by brief presentations of drifting sinusoidal gratings, shown monocularly or binocularly, in three human subjects. We found that responses to binocular stimuli are always significantly larger, with a binocular gain that can be as high as 10, and never lower than 2 (corresponding to linear summation), even at high contrast. This is the first demonstration of strong binocular summation at all contrasts. Using one-dimensional noise patterns we further demonstrated binocular summation.

Materials and methods

Subjects

Three human subjects (all male, two naive; ages 22–55) participated in the study. All subjects had normal or corrected-to-normal visual acuity and normal stereoacuity (evaluated using the Titmus, Randot, and Worth tests). Experimental protocols were approved by the Institutional Review Board concerned with the use of human subjects, and informed consent was obtained from all subjects. The study was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki); all personal identifiable information was handled in accordance with NIH privacy directives.

Apparatus

Subjects sat in a dark room, with their head stabilized using chin and forehead padded supports and a headband. Binocular stimulation was delivered using a custom built Wheatstone mirror stereoscope. The two CRT monitors (GDM-FW900; Sony, Tokyo, Japan) composing the stereoscope were viewed through 45° mirrors, arranged so that the optical distance of the monitors and the apparent distance of the binocular image seen through the mirrors were identical (521 mm). Each monitor screen covered 50° (H) by 32° (V) of visual angle, was set at a resolution of 1,280 columns by 800 rows, and a refresh rate of 140 Hz. Only the red channel was used, to minimize persistence (1 ms rise time, 4 ms fall time) and guarantee the absence of motion streaks (DeAngelis & Newsome, 2004). A single video card (GEForce GTX 580 Classified; EVGA, Brea, CA) was used to drive both monitors. The refresh timing of the two monitors was tightly synchronized, with the left eye image consistently preceding the right eye image by less than 50 μs. Luminance linearization was performed by interpolation following dense luminance sampling (using an LS100 luminance meter; Konica Minolta, Tokyo, Japan), independently for each monitor.

Horizontal and vertical positions of both eyes were recorded using an electromagnetic induction technique (Robinson, 1963). A scleral search coil embedded in a silastin ring (Skalar, the Netherlands; Collewijn, van der Mark, & Jansen, 1975) was placed in each eye following application of topical anesthetic (proparacaine HCl). Each coil output (sampled at 1,000 Hz) was calibrated at the beginning of each recording session by having the subject look at targets of known eccentricity. Peak-to-peak noise levels resulted in an uncertainty in eye position recording of less than 0.03°.

The experiment was controlled by two computers, one, running the Real-time EXperimentation software package (Hays, Richmond, & Optican, 1982), to manage the workflow and acquire and store the data, and the other, directly connected to the monitors, to generate the required visual stimuli in response to REX commands. This was accomplished using the Psychophysics Toolbox 3.0.8, a set of MATLAB (MathWorks, Natick, MA) scripts and functions (Brainard, 1997).

Behavioral paradigm: Ocular following

Trials were presented in blocks; each block contained one trial for each stimulus condition. All conditions within a block were randomly interleaved. Each trial began with the appearance of a binocular central fixation cross on a blank, mid-luminance (6.0 cd/m²), background. The subject was instructed to look at the center of the cross, and avoid making saccadic eye movements. After the subject maintained fixation within a small (1° on the side) invisible window around the fixation point for 800–1,100 ms, the fixation cross disappeared, and the visual stimulus sequence was presented for 170 ms. Subsequently the screen was blanked (again at mid-luminance), signaling the end of the trial. After a short intertrial interval, a new trial was started. If the subject blinked, or if saccades were
detected during the stimulus presentation epoch, the trial was discarded and repeated within the block.

**Behavioral paradigm: Perception**

Trials were presented in blocks; each block contained one trial for each stimulus condition. All conditions within a block were randomly interleaved. A two-interval forced choice (2IFC) paradigm with constant stimuli was employed. Each trial began with the appearance of a central fixation cross on a blank, mid-luminance (6.0 cd/m²), background. The subject was instructed to fixate the cross. After 800–1,100 ms, the fixation cross disappeared, and a first stimulus sequence (170 ms) was presented. At the end of presentation, a tone was emitted by a speaker and the screen was blanked (again at 6.0 cd/m²) for 500 ms. A second stimulus sequence (170 ms) was then presented. Finally, the screen was blanked (again at 6.0 cd/m²), and the subject had unlimited time to report, by pressing one of two buttons, whether the stimulus with higher apparent contrast was presented in the first or second interval. Once a response was recorded, a new trial was initiated.

**Visual stimuli**

All the stimuli had a mean luminance of 6.0 cd/m², and they were presented within a large aperture, centered on the screen. Outside the aperture the screen was blank, at mid-luminance. Two main classes of visual stimuli were used.

The first class of stimuli consisted of horizontal sinusoidal gratings, presented within a circular aperture (28° diameter). All gratings had a spatial frequency (SF) of 0.25 cpd, appeared suddenly and drifted upward or downward for 170 ms within the aperture, with a temporal frequency (TF) of 17.5 Hz. SF and TF were chosen, based on previous studies, to maximize the magnitude of the OFRs. The Michelson contrast of the grating was varied between 2.5% and 80% in one-octave increments. The stimulus was seen by either one eye only (“monocular” condition), while the other saw a mid-luminance blank screen, or by both eyes (“binocular” condition). In one subject (indicated as N3a), vertical stimuli drifting horizontally were also used. In that experiment, luminance was also higher (20.8 cd/m²) and all three guns of the CRT monitors were used.

The second class of patterns consisted of horizontal 1D random line stimuli (RLS), presented within a 28° diameter circular aperture. Each RLS was obtained by randomly assigning either a high or a low luminance value (symmetric around mean luminance) to each consecutive pair of rows of pixels (0.08°); it was moved at a speed of approximately 40°/s. Motion of the RLS was simulated by shifting either up or down (by an integer number of rows at each frame) a pattern larger than the screen behind the fixed aperture (i.e., the stimulus did not “wrap around”). The Michelson contrast, which is also equal to the root mean square (RMS) contrast for binary RLS, was varied between 2.5% and 80% in one octave increments. The two eyes could either see the same RLS (“binocular-same” condition), two different (uncorrelated) RLS (“binocular-different” condition), or the RLS could be presented to one eye while the other saw a mid-luminance blank screen (“monocular” condition).

In one subject we also ran an additional experiment in which drifting low-pass filtered horizontal RLS were presented. The gain of the filter was one below 0.375 cpd and zero above 0.75 cpd; the transition followed a raised-cosine function. The RMS contrast of the stimuli was varied between 1.5% and 24%. With non-binary RLS, RMS contrast is lower than, and limited by, Michelson contrast, but there is no fixed relationship between the two. We imposed a fixed value of RMS contrast (as opposed to Michelson contrast) because with noise stimuli RMS contrast has been shown to be a better indicator of stimulus strength (Moulden, Kingdom, & Gatley, 1990; Kukkonen, Rovamo, Tiippana, & Näätänen, 1993). The two eyes either saw the same RLS (“binocular-same” condition), two different (uncorrelated) RLS (“binocular-different” condition), or contrast-reversed (anti-correlated) versions of the same RLS (“binocular-opposite” condition).

**Statistical analysis**

All the measures reported here are based on eye velocity. The calibrated eye position traces (see Apparatus) were differentiated using a 21-point finite impulse response (FIR) acausal filter (47 Hz cutoff frequency). Trials with saccadic intrusions and unstable fixation that went undetected at run time were removed using an automatic procedure aimed at detecting outliers (Quaia, Optican, & Cumming, 2013). Average temporal profiles, time-locked to stimulus onset, were then computed over the remaining trials, separately for each stimulus condition. To remove the effect of components of the eye response related to the disengagement of fixation (Bostrom & Warzecha, 2010; Quaia, Sheliga, Fitzgibbon, & Optican, 2012), we report not the raw OFR, but rather the difference between the OFRs to upward and downward motion directions (rightward and leftward for N3a). The traces and measurements reported here are thus based on the difference between the average responses to stimuli.
containing motion energy in opposite directions. The strength of the OFR was quantified by computing, for each condition, the average difference eye speed in a time window that started at the latency of the response. The latency, which is a function of contrast, was defined as the time at which the difference eye speed became significantly ($p < 0.05$) different from zero, and was determined using a non-parametric bootstrap-based technique (Quaia, Sheliga, Optican, & Cumming, 2013). Latencies were initially computed separately for each stimulus condition. Inspection of the results revealed that monocular contrast is the prime determinant of latency. A lack of significant difference between the latency of OFRs to high contrast random dot stimuli presented monocularly or binocularly had also been previously reported in monkeys by others (Miles et al., 1986; Inoue et al., 2000). Because binocular responses are considerably larger, and thus provide a more reliable estimate of latency, we then used the latencies extracted from the binocular conditions (or “binocular-same” conditions for RLS stimuli) for the monocular responses as well (Figure 1). This has the added benefit of comparing monocular and binocular responses in identical time windows. The duration of the time window over which speed was averaged was equal to the latency of the response to the highest contrast stimulus (which varied between 67 and 75 ms across subjects). This ensured that the time window over which we measured the OFR was always restricted to the open-loop period (the period before motion of the eyes introduces a difference between motion on the display and motion on the retina that can affect the OFRs), and that measures from different contrast levels were computed over windows of the same duration.

All statistical analyses, including computations of standard errors and significance values, were carried out using non-parametric, bootstrap-based methods (Efron, 1982). A detailed description of the bootstrap procedures used can be found elsewhere (Quaia et al., 2012).

**Model fitting**

To quantify the relationship between the contrast of the stimulus and the strength of the OFR (the contrast response function, CRF), we measured the average eye speed in a time window, and used a Naka–Rushton function (also known as the Michaelis–Menten equation) to fit the CRF:

$$\mathcal{R}(c_s) = A \frac{c_s^n}{c_s^n + c_{50}^n} \quad (1)$$

where $c_s$ is the contrast of the stimulus, $A$ is a gain factor (which for the OFRs takes into account also the sensorimotor gain, and thus varies widely across subjects), $n$ determines the slope of the curve, and $c_{50}$ is the semi-saturation contrast (i.e., the contrast at which the response drops to $A/2$) and places the curve along the $c_s$ axis. Monocular and binocular conditions were fitted separately, with $c_s$ indicating monocular and binocular contrast, respectively.

Function fitting was performed using a simplex optimization algorithm in Python (using the numpy library). We did not minimize the summed squared error, but rather the $\chi^2$ measure:

$$\chi^2 = \sum_i \left( \frac{y_i - \hat{y}_i}{s_i} \right)^2$$

where, for each experimental condition $i$, we indicate...
with $y_i$ the mean OFR across trials, $s_i$ the standard error of $y_i$, and $\hat{y}_i$ the value of the fitted function for that condition. More consistent responses are thus weighted more heavily in determining the quality of a given fit, maximizing the likelihood of the model given the data.

The Naka–Rushton function provides descriptive fits, but offers no insights into how the differences between monocular and binocular OFRs might arise. We thus developed a simple cascade model, in which at the first stage each monocular input is passed through identical Naka–Rushton functions (representing monocular gain control mechanisms), whose outputs converge on a binocular stage, which is also modeled using a Naka–Rushton function (representing a binocular gain control mechanism). Because the gain parameters in the two stages are redundant, in the first stage we simply imposed a unitary gain. The basic model (illustrated in Figure 4A) has thus five parameters, and is described by the following equations:

$$y_{RE} = \frac{c_{RE}^n}{c_{RE}^n + c_{50}^n}$$
$$y_{LE} = \frac{c_{LE}^n}{c_{LE}^n + c_{50}^n}$$
$$y = y_{RE} + y_{LE}$$
$$z = G \frac{y_m}{y_m + y_{50}^m}$$  \(2\)

where $c_{RE}$ ($c_{LE}$) is the contrast of the stimulus presented to the right (left) eye. Model fitting was again carried out by minimizing $\chi^2$. The model was used only to fit the OFRs induced by sinusoidal drifting gratings. It was fitted separately to each subject’s data by using a simplex optimization algorithm in Python to find the set of five model parameters that resulted in the best match (i.e., minimal $\chi^2$) between the subject’s OFR and the model output $z$, across all monocular and binocular conditions at once. To compare the model performance to that of the Naka–Rushton fits one then needs to compare the $\chi^2$ for the model to the sum of the $\chi^2$ values for the Naka–Rushton fits to the monocular and binocular data.

### Results

**Binocular summation with sinusoidal gratings: OFRs**

We recorded the vertical eye movements induced by a drifting horizontal sinusoidal grating (SF = 0.25 cpd, TF = 17.5 Hz), which appeared suddenly on a mid-luminance background, and started drifting immediately (upward or downward, for 170 ms). The stimulus was presented either to only one eye (while the other eye saw a mid-luminance blank screen) or to both eyes, and its Michelson contrast was varied between 2.5% and 80%.

In Figure 1 we show the time course of the vertical version velocity (i.e., the mean vertical velocity of left and right eye) in response to monocular and binocular stimuli, in one subject. For monocular stimuli, the responses to the left-eye-only and right-eye-only stimuli were averaged. In the top row, responses are shown time-locked to the onset of the stimulus. In the bottom row the responses to 2.5%–40% contrast have been shifted back in time to bring them in alignment with the 80% contrast response. The time shifts were computed based on the latency of the binocular responses (first time when each response becomes significantly larger than zero), and then applied to both binocular and monocular responses. We chose to base our alignment on binocular responses because monocular responses were much smaller, making it difficult to reliably estimate their latency, especially at low contrasts. Note that the monocular responses are very well aligned (with the exception of the response to the 2.5% stimulus in this case). This held true for all subjects and in all experiments.

Four aspects of the data stand out: (1) The magnitude of responses increases with increasing contrast; (2) The latency of responses decreases with increasing contrast; (3) At each contrast level, binocular responses are larger than monocular responses; and (4) The binocular advantage is dramatic, e.g., a 5% binocular stimulus induces a stronger, albeit much delayed, OFR than an 80% monocular stimulus. The first two results were expected, and replicate previous observations (Miles et al., 1986; Gellman, Carl, & Miles, 1990; Sheliga, Chen, Fitzgibbon, & Miles, 2005; Miura et al., 2006; Barthelemy, Perrinet, Castet, & Masson, 2008). The relationship between monocular and binocular responses across contrast levels is the focus, and the novel contribution, of our study.

To quantify the strength of the response we computed the average eye speed of these time-shifted responses in a fixed time window (duration between 67 ms and 75 ms across subjects), indicated with a horizontal bar above the time axis in Figure 1. From here on we focus on this measure, which we simply call the OFR. The relationship between contrast and the OFR in our three subjects is shown in Figure 2, separately for monocular (blue) and binocular (orange) stimuli. In all subjects, responses to binocular stimuli were not only much larger than those to monocular stimuli of the same contrast, but also larger than what would be expected under linear summation (twice the monocular response, dashed blue line). Naka–Rushton functions (Equation 1) fitted to the data are also shown...
Because we reported the average speed of the two eyes, these results are also compatible with the hypothesis that each eye simply tracked the stimulus it saw (i.e., that under monocular conditions only one eye moved). We ruled this out by comparing the speed of the right and left eyes under monocular and binocular conditions (Figure 3). When the stimulus was presented only to the right eye (Figure 3A), the mean vertical (horizontal for N3a) speed (computed over the same time windows used for Figure 2) of the right (abscissa) and left (ordinate) eyes were very similar, for all subjects and across all contrasts tested. The same held true when stimuli were shown to the left eye only (Figure 3B). In both cases both eyes moved when only one eye saw the stimulus, and their mean speed was highly correlated (Pearson’s $r > 0.996$); there was a modest tendency for the eye seeing the stimulus to move slightly faster than the other eye, especially in subject N1. Under binocular stimulation (Figure 3C), there was again a very strong correlation between the speed of the two eyes (Pearson’s $r > 0.997$); in subjects N1 and N3 the right eye moved consistently faster, in N2 and N3a the left eye did, but again the differences were very small. Since the three panels share the same scale, it is also immediately apparent that each eye moved much faster under binocular than monocular conditions. The results cannot thus be explained by a disconjugate drive: Even under monocular stimulation, the OFRs were largely conjugate, with only a small vergence component, just like saccadic and smooth pursuit eye movements. This holds for both vertically and horizontally drifting stimuli (i.e., horizontal and vertical vergence appears to be equally stable during OFRs). This also indicates that recording the position of both eyes is not necessary to estimate the ratio between responses to monocular (averaging the movements induced by left eye only and right eye only presentations) and binocular stimuli of the same contrast. In Figure 3D we demonstrate this by computing this ratio based on the right eye speed only (abscissa) and on the version speed (ordinate). The two measures are highly correlated (Pearson’s $r > 0.99$), and never significantly different.

We just showed that the mean eye speed induced in the two eyes was highly correlated across stimuli, regardless of whether one or both eyes saw the target. As an aside, we wondered whether this tight correlation also held at the trial-by-trial level. We thus computed Pearson’s correlation between the mean speed in each eye in individual trials, over multiple presentations of the same stimulus. We found that it was always very
high, ranging (across subjects) between 0.84 and 0.89 when an 80% contrast binocular stimulus (i.e., the stimulus inducing the largest responses) was presented, and between 0.81 and 0.88 when a 2.5% monocular stimulus (i.e., the stimulus inducing the weakest responses) was presented. When all the trials collected, regardless of stimulus, were pooled together, the correlation varied between 0.80 and 0.84. Thus, even the trial-by-trial noise is highly correlated (i.e., conjugate) between the two eyes in OFRs. This implies that most of the trial-by-trial variability is a real property of the eye movements, not merely instrumental noise.

Since OFRs can be affected by the history of visual stimulation (Taki, Miura, Tabata, Hisa, & Kawano, 2009), we compared monocular and binocular responses separately on the subsets of trials that were preceded by higher and lower stimulus contrast. We found no significant differences in binocular gain (not shown), ruling out any effect of short-term adaptation on our results.

Table 1. Contrast tuning curves: Naka–Rushton fits.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Subject</th>
<th>Stimulus</th>
<th>A</th>
<th>n</th>
<th>c50</th>
<th>r²</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>N1</td>
<td>Monocular</td>
<td>0.78</td>
<td>1.27</td>
<td>9.19</td>
<td>19.11</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binocular</td>
<td>1.51</td>
<td>1.27</td>
<td>3.39</td>
<td>6.40</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>Monocular</td>
<td>0.50</td>
<td>2.05</td>
<td>4.99</td>
<td>6.18</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binocular</td>
<td>1.12</td>
<td>1.64</td>
<td>2.56</td>
<td>3.76</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>N3</td>
<td>Monocular</td>
<td>0.68</td>
<td>1.69</td>
<td>7.55</td>
<td>3.40</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binocular</td>
<td>1.54</td>
<td>2.46</td>
<td>4.78</td>
<td>1.64</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>N3a</td>
<td>Monocular</td>
<td>0.25</td>
<td>2.21</td>
<td>7.23</td>
<td>0.85</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binocular</td>
<td>0.62</td>
<td>2.08</td>
<td>4.52</td>
<td>2.17</td>
<td>0.994</td>
</tr>
<tr>
<td>6</td>
<td>N1</td>
<td>Monocular</td>
<td>0.78</td>
<td>1.06</td>
<td>20.44</td>
<td>12.60</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Same</td>
<td>1.10</td>
<td>1.56</td>
<td>3.61</td>
<td>10.60</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Diff</td>
<td>1.01</td>
<td>0.92</td>
<td>12.20</td>
<td>9.18</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>Monocular</td>
<td>0.44</td>
<td>2.36</td>
<td>12.87</td>
<td>1.41</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Same</td>
<td>0.75</td>
<td>3.27</td>
<td>4.54</td>
<td>9.04</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Diff</td>
<td>0.60</td>
<td>2.10</td>
<td>7.50</td>
<td>0.21</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>N3</td>
<td>Monocular</td>
<td>0.68</td>
<td>1.78</td>
<td>16.47</td>
<td>24.92</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Same</td>
<td>1.46</td>
<td>1.80</td>
<td>9.21</td>
<td>2.47</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Diff</td>
<td>1.20</td>
<td>1.38</td>
<td>20.31</td>
<td>2.71</td>
<td>0.997</td>
</tr>
<tr>
<td>7</td>
<td>N3</td>
<td>Bin-Same</td>
<td>1.82</td>
<td>1.92</td>
<td>2.50</td>
<td>5.72</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Diff</td>
<td>1.30</td>
<td>2.17</td>
<td>3.42</td>
<td>4.29</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bin-Opp</td>
<td>1.02</td>
<td>1.57</td>
<td>6.67</td>
<td>2.69</td>
<td>0.987</td>
</tr>
</tbody>
</table>

Figure 3. Monocular and binocular OFRs are conjugate. Whether the drifting grating was seen only by the right eye (A), only by the left eye (B), or by both eyes (C), the right (abscissa) and left (ordinate) eyes moved at very similar speeds, with binocular stimuli inducing stronger OFRs. The ratio between monocular and binocular responses (D) was the same whether it was computed based on the speed measured in one eye or the average speed of the two eyes. Both mean and SEM bars are shown. Type II regression lines are also shown.
Binocular summation with sinusoidal gratings: A simple model

To gain some insights into how such a strong enhancement for binocular stimuli might emerge, we fitted a simple two-stage model (Figure 4A) to the data. The contrast of the stimulus shown to each eye was fed to a first stage having a nonlinear input-output relationship described by a Naka–Rushton function. The outputs of the first stage for the left and right eyes were then summed. Finally, the resulting signal was fed to a second stage, also characterized by a Naka–Rushton input-output relationship (Equation 2). Model parameters were selected to best match the output of the second stage to the OFRs, across all monocular and binocular conditions at once.

The model fitted the data from each subject remarkably well (Figure 4B, one row per subject; parameter values and goodness-of-fit measures are listed in Table 2). The input to the first stage is monocular stimulus contrast. In the left column, we show the input–output relationship of the first stage (black); gray diamond symbols highlight the output of the first stage for the contrast levels that were tested in our experiment. Note that because a linear scale is used, the Naka–Rushton functions shown here appear more compressive than those in Figure 2, where a logarithmic scale was used for the abscissa, but this is deceiving (as can be inferred by comparing values for \( n \) and \( c_50 \) in Tables 1 and 2). In the right column we show the input–output relationship of the second stage (black). The input to the second stage \( (y) \) is the sum of the outputs of the two first (monocular) stages (Equation 2), and it thus equals the location along the ordinate of the diamonds in the left column for monocular stimuli (blue), and twice that value for binocular stimuli (orange). The position along the ordinate of the symbols is equal to the OFRs shown in Figure 2.

The input–output relationships of the first and second stage are consistent across subjects. The first stage is compressive, with the output saturating for contrasts above 20%; the second stage is instead expansive at low and intermediate inputs, and only slightly compressive at large inputs. These properties are compatible with the strong contrast normalization observed in magnocellular neurons in the retina and LGN (Derrington & Lennie, 1984; Kaplan & Shapley, 1986; Sclar, Maunsell, & Lennie, 1990; Benardete,) input (sum of the outputs of the two monocular stages) is shown, together with OFRs for monocular (blue) and binocular (orange) stimuli (both mean and SEM bars are shown). Black lines show the input-output Naka-Rushton relationship of the two stages. Parameter values and goodness-of-fit measures are listed in Table 2.
predict that measured experimentally in the LGN (model could then be used to predict values of with the LGN, and the second stage with cortex, the linearities (Albrecht & Geisler, 1991). by virtue of their (typically expansive) output non-linearities (Albrecht & Geisler, 1991).

If we were to identify the first stage of our model with the LGN, and the second stage with cortex, the model could then be used to predict values of $c_{50}$ measured experimentally in the LGN ($c_{50}$ in Table 2), and in cortex (given the high similarity of the model and Naka–Rushton fits to the data, approximately equal to $c_{50}$ in Table 1). Notably, our model would then predict that $c_{50}$ in visual cortex is different for monocular and binocular stimuli. For binocular stimuli, our model predicts that $c_{50}$ should be very similar when measured at the output of the LGN and cortex; with monocular stimuli it is instead predicted to be higher in cortex. However, magnocellular neurons in cortex are reported to have a lower $c_{50}$ than magnocellular neurons in the LGN (Sclar et al., 1990), indicating that caution must be taken in making parallels between the model stages and neural structures. Nonetheless, we will later show that OFRs might be mediated by a small subset of cortical neurons, and thus direct comparisons with population averages might not be appropriate (see Discussion).

Table 2. Contrast tuning curves: Two-stage model fits.

<table>
<thead>
<tr>
<th>Subject</th>
<th>$n$</th>
<th>$c_{50}$</th>
<th>$G$</th>
<th>$m$</th>
<th>$y_{50}$</th>
<th>$\chi^2$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>1.00</td>
<td>3.51</td>
<td>1.77</td>
<td>2.82</td>
<td>1.10</td>
<td>25.14</td>
<td>0.993</td>
</tr>
<tr>
<td>N2</td>
<td>1.35</td>
<td>2.48</td>
<td>1.33</td>
<td>2.97</td>
<td>1.16</td>
<td>10.08</td>
<td>0.995</td>
</tr>
<tr>
<td>N3</td>
<td>1.51</td>
<td>4.21</td>
<td>2.34</td>
<td>2.37</td>
<td>1.46</td>
<td>8.34</td>
<td>0.997</td>
</tr>
<tr>
<td>N3a</td>
<td>1.41</td>
<td>3.75</td>
<td>0.83</td>
<td>2.76</td>
<td>1.33</td>
<td>2.98</td>
<td>0.997</td>
</tr>
</tbody>
</table>

Kaplan, & Knight, 1992; Kremers, Silveira, & Kilavik, 2001; Solomon, White, & Martin, 2002; Webb et al., 2002; Priebe & Ferster, 2006; Solomon, Lee, & Sun, 2006; Alitto & Usrey, 2008; Camp, Tailby, & Solomon, 2009; Alitto, Moore, Rathbun, & Usrey, 2011), where neurons are monocular, and with the proposal that 2002; Priebe & Ferster, 2006; Solomon, Lee, & Sun, 2001; Solomon, White, & Martin, 2002; Webb et al., 2001; Ding & Sperling, 2006; Meese et al., 2006; Baker & Movshon, 1989; Cagenello et al., 1993; Zlatkova et al., 2001; Ding & Sperling, 2006; Meese et al., 2006; Baker et al., 2007; Hess et al., 2007; Pineles et al., 2014). In stark contrast, in our subjects binocular stimuli induced OFRs that were at least twice as large as those induced by monocular stimuli, regardless of contrast. Inspection of Figure 2 reveals that a 80% contrast monocular stimulus induced a weaker OFR than a 10% binocular stimulus (which is quite dim). It would thus seem that perception and ocular following obey completely different rules in terms of binocular summation. There is however a potential flaw in this line of reasoning: The stimuli used for perceptual experiments are very different from the ones we used to record OFRs. In typical psychophysical experiments, the stimuli are much smaller, of much higher SF, static, and are usually present on the screen for longer times. Any of these differences might in principle explain the discrepancy.

To test this hypothesis, we had our three subjects perform a contrast matching psychophysical experiment, using the same stimuli that were used for the eye movement recordings. The experiment was in the form of a two-interval forced choice (2IFC): Two stimuli were presented in succession for 170 ms each, separated by a short interval during which the screen was blank. Subjects had then to report which interval contained the stimulus with higher perceived contrast. During one
of the two intervals a 10% contrast binocular grating (reference stimulus) was shown; during the other interval a grating with one of four contrast levels (0, 2.5, 5, or 7.5%) was shown to one eye (eye 1, which in different trials could be either the left or the right eye) and an identical grating with one of six contrast levels (5, 10, 12.6, 15.88, 20, or 40%) was shown to the other eye (eye 2). In Figure 5A we plot, for the three subjects, the percentage of times that the reference stimulus was reported to have lower contrast. A separate psychometric curve was fitted for each of the four contrast levels in eye 1. Slight left/right eye asymmetries were averaged by pooling the data regardless of whether eye 1 was the left or the right eye in a given trial. As expected, when the contrast shown to eye 1 decreased, the point at which the corresponding psychometric curve crossed 50% (i.e., the contrast that needed to be shown to eye 2 to match the perceived contrast of the reference stimulus) increased. However, even with monocular stimuli (zero eye 1 contrast), this point never reached 20%.

In Figure 5B this same data are plotted in a different format, one which is routinely used in the psychophysical literature. Here points that have the same apparent contrast, based on the fits shown in Figure 2A, are shown, together with a fit using the power summation rule (Legge & Rubin, 1981):

\[ 2c_{BE}^p = c_{RE}^p + c_{LE}^p \]

This equation has a single parameter, the exponent \( p \), and in our experiments \( c_{BE} \) was always equal to 10%. Three values of \( p \) are particularly meaningful. When \( p = 1 \) (diagonal dashed line) the perceived contrast is equal to the linear sum of the contrasts from the two eyes; when \( p = 2 \) (dashed circle) perceived contrast is determined by the Pythagorean sum of the contrasts from the two eyes (quadratic summation); finally, when \( p = \infty \) (dashed square at 10% contrast), perceived contrast is equal to the highest contrast in the two eyes. In our three subjects, the best fitting \( p \) (shown in each panel) is between 1.5 and 2. This is within the range of values reported in previous psychophysical studies.
Importantly, this experiment allowed us to determine monocular and binocular stimuli that were perceived as having the same contrast. The binocular stimulus always had 10% contrast in each eye (the reference stimulus), whereas the perceptually equivalent monocular stimulus varied across subjects: It was 17.2% in N1, 14.2% in N2, and 16.9% in N3. Armed with these results we can now go back to our eye movement recordings, and find the magnitude of the OFRs corresponding to these stimuli, which are indicated with black square symbols in Figure 2. These stimuli obviously induced vastly different OFRs, despite being perceptually equivalent. The ratio between the OFRs induced by the perceptually equivalent binocular and monocular stimuli is 2.24 in N1, 2.25 in N2, and 2.44 in N3. It thus appears that perception and ocular following rely on at least partially distinct neuronal substrates, which combine binocular inputs according to different rules.

**Binocular summation with 1D noise gratings: OFRs**

In the binocular conditions tested above, the eyes saw two identical stimuli. It has been reported that when the two eyes see sinusoidal gratings that have different phases, they are perceived as having lower contrast than when they have the same phase, especially when the contrast of the stimuli is low and the phase difference is large (Westendorf & Fox, 1974; Bacon, 1976; Blake & Rush, 1980; Green & Blake, 1981; Baker, Wallis, Georgeson, & Meese, 2012). This finding might be related to the observation that the response of most binocular neurons in visual cortex is sensitive to the interocular phase of the binocular inputs (Poggio & Fischer, 1977; Ohzawa & Freeman, 1986; Gonzalez & Perez, 1998; Cumming & DeAngelis, 2001). It would thus be interesting to test how interocular phase differences affect the OFRs. It has already been shown that applying a disparity to a stimulus causes a modest attenuation of the OFRs (Masson, Busettini, Yang, & Miles, 2001), but the presence of a disparity also triggers a short latency disparity verge response (Busettini, Masson, & Miles, 1996; Busettini, Miles, & Krauzlis, 1996), potentially contaminating the measure. By using noise stimuli it is however possible to manipulate interocular correlation between the two eyes without inducing significant vergence responses. This can be achieved by simply presenting different (i.e., uncorrelated) noise stimuli to the two eyes.

We presented random line stimuli (RLS) spanning a range of RMS contrasts to our three subjects. The stimuli always drifted up or down at high speed, and were presented either to only one eye or to both eyes. In binocular presentations the two eyes saw either the same stimulus, or two different RLS (drifting at the same speed and in the same direction). As we did before, we identified the open-loop time window for each subject, and used the latencies of the OFRs elicited in the binocular-same conditions to time-shift the responses as a function of contrast. The relationship between contrast and the OFR in our three subjects is shown in Figure 6, separately for monocular (blue),
neurons tuned to zero disparity, as proposed by others to account for the effect of interocular phase in perceptual binocular summation (Baker et al., 2012). Such neurons would produce maximal responses for binocular-same stimuli and weaker responses for binocular-different (uncorrelated) stimuli (Cumming & Parker, 1997), just as we observed.

Because neurons tuned to zero disparity typically produce (Cumming & Parker, 1997) even weaker responses to anticorrelated stimuli (i.e., stimuli obtained by presenting contrast-reversed, but otherwise identical, images to the two eyes), we further tested this hypothesis by presenting anti-correlated stimuli to one of our subjects. In this experiment low-pass filtered RLS stimuli were used, and three interocular binocular correlation conditions were tested: binocular-same, binocular-different (i.e., uncorrelated), and binocular-opposite (i.e., anticorrelated). We found (Figure 7A; parameter values and goodness-of-fit measures for the fits are listed in Table 1) that once again binocular-different stimuli (gray) induced weaker responses than binocular-same stimuli (orange), at all contrasts; importantly, binocular-opposite stimuli (green) induced even weaker OFRs. At high-contrast, the ratio between binocular-same and binocular-opposite OFRs was slightly larger than 2, indicating that binocular-opposite stimuli were probably as effective as monocular stimuli (which were not included in this experiment).

This last observation suggests a potential alternative interpretation for our results: Responses to stimuli that are different in the two eyes might be weaker as a result of binocular rivalry (Blake & Wilson, 2011), which might suppress the input from one eye. Rivalry might be expected to be proportional to the difference between the images presented to the two eyes, and thus could be stronger for binocular-opposite stimuli (in which all lines have opposite contrast) than for binocular-different stimuli (in which, on average, only half of the lines are different). While intuitively attractive, this explanation does not stand up to scrutiny. First, the stimuli used in this experiment did not appear perceptually rivalrous. This is not surprising, given that the RLS presented to the two eyes had the same orientation, spatial frequency content, and drifted at the same speed. Second, it is well-known that rivalry takes time to ensue. Rivalry onset is usually estimated to take anywhere between 150 and 450 ms from stimulus onset (Wolfe, 1983; Blake, 1989), although recording OFRs to stimuli that differ in the two eyes in terms of both orientation and temporal frequency content, and are perceptually rivalrous, we have reported that the first signs of rivalry could actually appear as early as 110 ms from stimulus onset (Quaia, Optican, & Cumming, 2016). If we examine the time course of the OFRs induced by 24% RMS contrast
stimuli in our experiment (Figure 7B) we find however that the three classes of stimuli induced different responses from the very beginning. For example, binocular-same stimuli induced a response that became significantly different from zero ($p < 0.05$) 73 ms after stimulus onset, and was already significantly larger than that induced by binocular-opposite stimuli only ten ms later (83 ms from stimulus onset). This is at least 30 ms faster than any rivalry onset estimate we know of.

**Discussion**

We characterized binocular summation (the response boost conferred by binocular over monocular stimulation) for ocular following eye movements, across a wide range of supra-threshold contrast levels. We discovered that, for OFRs, binocular summation is large at all contrasts when the two eyes are presented with identical stimuli, but is smaller when uncorrelated (or anticorrelated) stimuli are shown to the two eyes. Differences between monocular and binocular OFRs to high contrast drifting random-dot stimuli had been reported before in monkeys, but only qualitatively and without a characterization of the sensitivity of binocular summation to contrast and interocular phase differences (Miles et al., 1986; Inoue et al., 2000). A modest dependency of the OFR on the absolute disparity of a high-contrast noise stimulus has been previously reported (Masson et al., 2001), but those results were potentially contaminated by short latency vergence responses.

As noted in the Introduction, sizable binocular summation for perception has been observed only at very low contrasts, at or just above threshold. Binocular summation is thus quite different for perception and OFRs, adding it to the list of phenomena that exhibit different properties when evaluated using perception and eye movements (Bostrom & Warzecha, 2010; Simoncini, Perrinet, Montagnini, Mamassian, & Masson, 2012; Blum & Price, 2014; Glasser & Tadin, 2014; Price & Blum, 2014; Quaia et al., 2016). It is generally difficult to pinpoint the source of such differences, especially when perception and eye movements are measured using stimuli that differ in their spatiotemporal content. However, by directly comparing OFRs to monocular and binocular stimuli having identical perceived contrast (Figure 2), we ruled out this source of uncertainty. A possible explanation for this large difference is that OFRs might rely only on a small subset of V1 neurons, which might not be representative of the neurons involved in perception. First, only direction-selective (DS) neurons, which are approximately 20% of all V1 neurons (De Valois, Yund, & Hepler, 1982; Orban, Kennedy, & Bullier, 1986; Hawken, Parker, & Lund, 1988; Gur, Kagan, & Snodderly, 2005), can contribute to a directional response. Second, the OFR is tuned to low spatial frequencies and high temporal frequencies (i.e., high speeds), which strongly activate only a subset of DS neurons in V1 (mostly part of the magnocellular pathway). Finally, our results with noise stimuli indicate that OFRs might be dominated by neurons tuned to zero or near-zero disparity, again a subset of V1 neurons. This might explain why recording methods that are sensitive to the pooled activity of large populations of V1 neurons, such as fMRI (Moradi & Heeger, 2009) and EEG (Baker & Wade, 2017), find limited binocular summation, more in line with perceptual reports than with OFRs. Both these measures and perceptual contrast judgments might be dominated by parvocellular signals, with little regard for direction selectivity, speed preference, and preferred disparity. This is sensible, since it is reasonable to expect that, when carrying out perceptual contrast matching, a subject would rely mostly on parvocellular neurons, which encode contrast linearly, instead of magnocellular signals, which quickly saturate with contrast (Derrington & Lennie, 1984; Kaplan & Shapley, 1986; Sclar et al., 1990; Usrey & Reid, 2000; Alitto et al., 2011). OFRs instead are known to be mediated by neurons in areas MT and MST (Takemura et al., 2007; Masson & Perrinet, 2012), whose activity is in turn determined mostly by magnocellular LGN neurons (Maunsell, Nealey, & DePriest, 1990), and accordingly saturate quickly with contrast.

A simple two-stage model, compatible with what is known about the dorsal stream of visual processing, accounted very well for the OFRs induced by monocular and binocular sinusoidal gratings (Figure 4). Extending it by making its second stage tuned to zero-disparity would allow it to also account for the effect of inter-ocular phase differences seen with our noise stimuli (Figures 6 and 7). One limit of this explanation is that, whereas neurons tuned to zero-disparity are present in significant numbers in cortical area V1 (Cumming & DeAngelis, 2001; Prince, Cumming, & Parker, 2002), they appear to be conspicuously rare in extrastriate visual areas MT and MST (DeAngelis & Newsome, 1999; Takemura, Inoue, Kawano, Quaia, & Miles, 2001; Cumming & DeAngelis, 2001), which are known to mediate OFRs (Takemura et al., 2007). One possible explanation is that the disparity tuning to flashed and drifting stimuli might be different, but the available evidence argues against this interpretation (Anzai, Ohzawa, & Freeman, 2001; Palanca & DeAngelis, 2003). Alternatively, we can only speculate that the region of MST from which these neurons have been recorded is not the same as the one that drives OFRs, where perhaps zero-disparity tuned neurons might be found. Since our
results indicate that only neurons that respond more strongly to binocular-same than to binocular-different (or binocular-opposite) stimuli are good candidates for driving OFRs, this is obviously an important subject for future electrophysiological studies.

Acknowledgments

This research was supported by the Intramural Research Program of the National Eye Institute, NIH, DHHS. We thank Boris Sheliga for assistance in running the experiments.

Commercial relationships: none.
Corresponding author: Christian Quaia.
Email: quaiac@nei.nih.gov.
Address: Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA.

References


Camp, A., Tailby, C., & Solomon, S. (2009). Adaptable mechanisms that regulate the contrast response of


Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: SIAM.


Quaia, C., Sheliga, B., Fitzgibbon, E., & Optican, L. (2012). Ocular following in humans: Spatial prop-


