Temporal Evolution of Pattern Disparity Processing in Humans

Christian Quaia, Boris M. Sheliga, Lance M. Optican, and Bruce G. Cumming

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892

Stereo matching, i.e., the matching by the visual system of corresponding parts of the images seen by the two eyes, is inherently a 2D problem. To gain insights into how this operation is carried out by the visual system, we measured, in human subjects, the reflexive vergence eye movements elicited by the sudden presentation of stereo plaids. We found compelling evidence that the 2D pattern disparity is computed by combining disparities first extracted within orientation selective channels. This neural computation takes 10–15 ms, and is carried out even when subjects perceive not a single plaid but rather two gratings in different depth planes (transparency). However, we found that 1D disparities are not always effectively combined: When spatial frequency and contrast of the gratings are sufficiently different pattern disparity is not computed, a result that cannot be simply attributed to the transparency of such stimuli. Based on our results, we propose that a narrow-band implementation of the IOC (Intersection of Constraints) rule (Fennema and Thompson, 1979; Adelson and Movshon, 1982), preceded by cross-orientation suppression, underlies the extraction of pattern disparity.

Introduction

Binocular vision endows humans with the ability to infer the 3D organization of objects in the environment. However, this carries a cost: corresponding features in the images on the two retinæ must be correctly matched (Julesz, 1971). Although horizontal disparities play the major role in depth perception, vertical disparities are also important (Helmholtz, 1925). Consequently, the correspondence problem must be solved in two dimensions: a 2D disparity vector must be extracted. Two strategies could be used to perform this computation. First, a direct matching in 2D (e.g., a 2D cross-correlation between the two retinal images, or the matching of monocular 2D features) could be carried out. Alternatively, the 2D problem could be decomposed into the extraction of disparity along various directions, which could then be combined to yield a 2D vector, analogous to the process proposed for pattern motion perception (Fennema and Thompson, 1979; Adelson and Movshon, 1982). While the first option is conceptually simpler, the pervasiveness of orientation-selective disparity detectors in the visual system (Cumming and DeAngelis, 2001) suggests a neural implementation of the alternative.

Based on psychophysical observations using stereo “plaids” (the sum of two sinusoidal luminance gratings having different orientations), Farell (1998) concluded that pattern disparity does

indeed drive depth perception in human observers. In that same study, he also used disparity adaptation to argue that the computation of pattern disparity is preceded by, and relies on, the extraction of the disparity of the individual gratings (i.e., within orientation-selective channels). However, interpreting the effects of adaptation depends on assumptions about the site and mechanism underlying disparity adaptation, leaving room for alternative interpretations (Delicato and Qian, 2005).

To sidestep the limits of depth perception, we have taken advantage of the observation that when an image is suddenly presented binocularly, the disparity in the image induces a reflexive short-latency vergence eye movement (Busettini et al., 1996). This vergence movement, termed the disparity vergence response (DVR), is induced by vertical, horizontal, and oblique disparities (Rambold and Miles, 2008), and so is well suited for studying how 2D disparities are computed. Furthermore, the evolution of the DVR can be finely resolved in time, providing a unique window into the time course of the neural computations underlying disparity detection.

We measured DVRs to plaids composed by summing two stimuli varying along one spatial dimension (1D gratings) but having different orientations. We chose the orientation and disparity of each grating so as to dissociate the direction of component and pattern disparity, enabling us to analyze the time evolution of disparity detection. We found that component disparity is extracted first, and that pattern disparity emerges only 10–15 ms later. Furthermore, manipulating the component spatial frequencies and contrasts revealed that the pattern disparity computation is quite sensitive to these parameters, and it is not based on a 2D cross-correlation. Our results decisively support, and significantly extend, the hypothesis, previously put forward by Farell (1998), that pattern disparity is computed on the basis of 1D component disparities.

Received Sept. 10, 2012; revised Nov. 21, 2012; accepted Dec. 27, 2012.

Author contributions: C.Q., B.M.S., L.M.O., and B.G.C. designed research; C.Q. and B.M.S. performed research; C.Q. analyzed data; C.Q., L.M.O., and B.G.C. wrote the paper.

This research was supported by the Intramural Research Program of the National Eye Institute, NIH, Department of Health and Human Services.

The authors declare no competing financial interests.

Correspondence should be addressed to Christian Quaia, Laboratory of Sensorimotor Research, National Eye Institute, NIH, Department of Health and Human Services, 49 Convent Drive, Room 2A50, Bethesda, MD 20892. E-mail: quaia@nei.nih.gov.

DOI:10.1523/JNEUROSCI.4318-12.2013

Copyright © 2013 the authors 0270-6474/13/333465-12\$15.00/0

Figure 1. Visual stimuli. We used different types of plaids, and here monocular samples from each type are reproduced. **A**, Sinusoidal plaids. **B**, Square wave plaids. **C**, 1D binary random noise line plaids. **D**, All the binocular stimuli used in our experiments were obtained by summing a cardinal (vertical or horizontal) grating to an oblique grating. Here the gratings seen by each eye are indicated with vertical and oblique lines (black for the plaid seen by the right eye, gray for that seen by the left eye), which can be thought of as tracing a line of constant luminance in a sinusoidal grating (e.g., the maximum or minimum luminance). In our experiment the cardinal grating was shifted in the two eyes (CD), whereas the oblique grating was not (hence the oblique gray and black lines are superimposed). With this particular choice of plaid (which we term unidispary plaid) the PD vector (outlined arrow) is parallel to the orientation of the zero-disparity grating. The PD vector is equal to the vector by which the right eye image (black lines) must be shifted to match the left eye image (gray lines), and it is also equal to the vector between the intersection of the black (right eye) lines (black bull's eye) and the intersection of the gray (left eye) lines (gray bull's eye).

Materials and Methods

Subjects. Three male subjects participated in the experiments; two were authors and one was unaware of the experimental questions being investigated. All had normal or corrected-to-normal visual acuity and normal stereoacuity. Experimental protocols were approved by the Institutional Review Board concerned with the use of human subjects.

Visual apparatus. The subjects sat in a dark room, and were positioned so that their eyes were located approximately in the center of a cubic box (70 cm side) containing orthogonal magnetic-field generating coils. Their chin and forehead rested on padded supports, and their head was stabilized using a head band. Visual stimuli were presented dichoptically using a Wheatstone mirror stereoscope. Each eye saw a CRT monitor (Sony GDM-F520) through a 45° mirror, creating a binocular image straight ahead at a distance of 521 mm from the corneal vertex, which was also the optical distance to the two monitor screens. Each monitor screen covered 42° (horizontal) by 32° (vertical) of visual angle. The background luminance was set to 20.8 cd/m² (as reported by a Konica Minolta LS100 luminance meter). The RGB signals from the video card (NVIDIA Quadro FX 5600) provided the inputs to an attenuator whose output was connected to a single channel of a video signal splitter (Black Box Corp., AC085A-R2); the video outputs of the splitter were then connected to the RGB inputs of the monitor. In this way only grayscale images could be presented, but with a higher luminance resolution (12 bits) than normally possible (8 bits). Luminance linearization was performed by interpolation following dense luminance sampling.

Eye movement recording. A scleral search coil embedded in a silastin ring (Skalar) (Collewijn et al., 1975) was placed in each of the subject's eyes following application of topical anesthetic (proparacaine HCl). The horizontal and vertical orientations of the eyes were recorded using an electromagnetic induction technique (Robinson, 1963). These outputs were 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°. Coil signals were sampled at 1000 Hz.

Experiment control. The experiment was controlled by three computers, communicating over an Ethernet with TCP/IP. The Real-time Experimentation software package (Hays et al., 1982), running on the master computer under the QNX operating system, was responsible for providing the overall experimental control as well as acquiring, displaying, and storing the eye movement data. The other two machines, directly connected to the CRT displays, run under the Windows XP operating system and generated the required visual stimuli in response to REX commands. This was accomplished using the Psychophysics Toolbox v. 3.0.8, a set of Matlab (MathWorks) scripts and functions (Brainard, 1997). Frame synchronization between the two monitors was enforced by frame-locking each computer's video card (NVIDIA Quadro FX 5600) using NVIDIA Quadro G-Sync cards.

Behavioral paradigm. 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 central fixation cross (width 10°, height 5°, thickness 0.05°) on a gray background (20.8 cd/m²). 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 800–1100 ms, the fixation cross disappeared, and the visual stimuli appeared simultaneously on both screens. After 200 ms both screens turned gray (again at 20.8 cd/m²), 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. With few exceptions, a single experiment required multiple daily recording sessions to collect enough trials from a subject (we collected between 150 and 450 trials for each condition, depending on the signal-to-noise ratio, and between 1000 and 2400 trials in a session; the number of conditions varied across experiments).

Visual stimuli. The basic visual stimuli for our experiments were plaids obtained by summing two sinusoidal 1D gratings having different orientations (Fig. 1A). The stimuli filled the entire screen. The average luminance of each grating and of the plaid was equal to 20.8 cd/m², and the plaid stimuli were obtained by adding, pixel by pixel, to this mean luminance the luminance deviations of each grating from the mean luminance. Unless otherwise noted, the Michelson contrast (Michelson, 1927) of each grating was 32%. In most cases the spatial frequency (SF) used was 0.25 cpd, as previous experiments (Sheliga et al., 2006) indicated that most subjects respond strongly to this SF. Exceptions are indicated in the text. The component disparity (measured in the direction orthogonal to the grating orientation) was controlled separately for each grating, and was either zero or a quarter of the stimulus wavelength (90° phase disparity). We call plaids in which one of the two gratings has zero disparity unidispary plaids, in analogy to the unikinetic plaids used in motion experiments (Gorea and Lorenceau, 1991; Dobkins et al., 1998).

In a second experiment we used instead plaids obtained by summing two square wave gratings (Fig. 1B). The average luminance of each grating and of the plaid was equal to 20.8 cd/m²; the high and low luminances of each grating were equal to the maximum and minimum luminances used for the 32% contrast sinusoidal gratings. The Michelson contrast was thus the same as that of the sinusoidal gratings, but the root mean square contrast of the square wave stimuli was larger. Once again the SF of each grating was 0.25 cpd and the phase disparity of each component was either 0° or 90°. Each pixel in the plaid stimulus could thus have one of three luminance values: the mean luminance, the mean luminance plus the difference between the high and low luminance in a grating, or the mean luminance minus the difference between the high and low luminance in a grating.

Figure 2. Measuring DVRs. DVRs to two stimuli having opposite disparity are often asymmetric. **A**, This subject (BMS) when presented with a vertical sinusoidal grating (SF = 0.25 cpd, contrast = 32%, $\pm 90^\circ$ phase disparity) produced a larger DVR for an uncrossed stimulus (dashed white on dark gray) than for a crossed one (dashed black on light gray). The time course of the response was, however, very similar. To improve the signal-to-noise ratio, and to eliminate a component of the response that is not strictly related to the disparity of the stimulus, we thus subtracted the two (white on black). Both mean (thin lines) and $\pm$ SEM (thick shades) of the response are plotted. Two different methods were used to determine the latency of response (see Statistical analysis in Materials and Methods). They always yielded similar values; to avoid clutter in the remaining figures only the L_5 measure is indicated. **B**, The same subject, when presented with a horizontal sinusoidal grating responded more vigorously to a left-hyper (dashed black on light gray) than to a right-hyper (dashed white on dark gray) disparity. Again, the difference between the two responses (white on black) is used throughout our study.

Finally, in a third experiment we used plaids obtained by summing two 1D binary random noise line patterns (Fig. 1C). In this case, each four-pixel (0.1°) wide line of pixels in a pattern was randomly assigned either a high or a low luminance value (equal to the maximum and minimum luminance used for sinusoidal and square wave gratings). The average luminance of each line pattern and of the plaid was equal to 20.8 cd/m^2 . Only purely horizontal, vertical, or $\pm 45^\circ$ orientations were used. The position disparity of each component was either 0° or 0.2° . In this experiment the stimuli were presented in a circular aperture (32° diameter); the surround was gray (20.8 cd/m^2), and the aperture had zero disparity.

Note that in unidispersion plaids the pattern disparity vector (i.e., the rigid translation necessary to overlap the images seen by the left and right eyes) is parallel to the orientation of the zero disparity grating. For example, in Figure 1D we represent a plaid composed by summing a vertical and an oblique grating. Each line indicates a grating “feature” (such as a line of constant luminance in a sinusoidal grating, or an edge in a square wave grating or noise 1D pattern); black lines indicate the gratings seen by the right eye (R), whereas gray lines indicate the gratings seen by the left eye (L). In this example, the vertical grating has disparity, whereas the oblique grating has zero disparity. The only component disparity (CD) in this plaid is thus horizontal, as indicated by the black arrow. However, the pattern disparity (PD, outlined arrow) is parallel to the oblique grating, and thus has a vertical component (even though neither grating has vertical disparity). The pattern disparity vector can be determined by noting that this is the vector by which the black (right eye) grating lines need to be rigidly shifted to overlap the gray (left eye) grating lines. Alternatively, one can use the Intersection of Constraints (IOC) rule, based on which the pattern disparity vector is equal to the vector from the intersection of the black grating lines (black bull’s eye) to the intersection of the gray grating lines (gray bull’s eye).

Data analysis. All the measures reported herein are based on vergence velocity. The calibrated eye position traces (see Eye movement recording, above) were differentiated using a 21-point FIR acausal filter (47 Hz cutoff frequency). The difference between the horizontal (vertical) velocity of the left and right eyes was then computed, yielding the horizontal (vertical) vergence velocity for each trial. Trials with saccadic intrusions and unstable fixation that went undetected at run time were removed through an automatic procedure aimed at detecting outliers. For each velocity signal (left eye horizontal, right eye horizontal, left eye vertical, right eye vertical, horizontal vergence, vertical vergence) at each time point (0–199 ms from stimulus onset, in 1 ms increments), trials for which the velocity deviated more than ± 4.5 SDs from the mean (across

all the valid trials for a given condition) were excluded. This was repeated iteratively until no trials were excluded. The fraction of trials so excluded varied from session to session, and usually increased during a session. Typically, during the first 20 min within a session no more than 5% of the trials were excluded, but usually by the end of the session this fraction increased, sometimes to as much as 20%. Eye discomfort due to anesthetic effect wearing off and decrease in concentration are the most likely causes. Average temporal profiles, time-locked to stimulus onset, were then computed over the remaining trials, separately for each stimulus condition.

Opposite stimulus disparities (e.g., crossed and uncrossed by the same magnitude) are known to elicit DVRs that are not simply opposites (Busetini et al., 1996). The most striking example of this asymmetry can be seen when very large disparities are applied: Very large positive and negative disparities induce the same (non-zero, and different across subjects) DVR. This response, independent of the sign of the stimulus disparity, is called a default response, and is probably partly related to the disengagement of fixation. In Figure 2 we show, for one subject, the DVRs induced by

a single grating, either oriented vertically with horizontal disparity (crossed or uncrossed, Fig. 2A), or oriented horizontally with vertical disparity (left- or right-hyper, Fig. 2B). Obviously the DVRs are in the direction that tends to reduce the disparity of the stimuli, but there are asymmetries between the responses to opposite disparity. To remove this spurious component, and to increase the signal-to-noise ratio, most disparity vergence studies (Sheliga et al., 2006, 2007; Miura et al., 2008; Rambold and Miles, 2008; Rambold et al., 2010) report not the raw DVR, but rather the difference between the DVRs to opposite disparities. We did so here as well: The difference trace (Fig. 2) provides a more faithful representation of the DVR imputable to the disparity signal alone. The traces and measurements reported here are thus based on the difference between the average response to a crossed (left-hyper) disparity stimulus and that to the same size uncrossed (right-hyper) disparity stimulus.

Statistical analysis. We used bootstrap-based methods (Efron, 1982) for all our statistical analyses. Here, we briefly describe the procedures used.

To evaluate the DVRs statistically we took, for each stimulus type, the N_X trials (T_{Xi} being the vergence speed measured during these trials) associated with crossed (left-hyper) disparity and the N_U trials (T_{Uj}) associated with uncrossed (right-hyper) disparities. We then defined the difference between the average vergence speed for crossed and uncrossed trials $V = \sum_i T_{Xi}/N_X - \sum_j T_{Uj}/N_U$ as the mean vergence speed in response to the stimulus type. To obtain a confidence interval around this measure, we then selected, with replacement, N_X trials from T_{Xi} and N_U trials from T_{Uj} , and computed the difference of their means. This was repeated 1000 times, producing a distribution of bootstrapped mean average speeds V^* . Sixty-eight percent confidence intervals ($\pm$ SEM for Gaussian distributions) for the mean vergence speed V could then be obtained by computing, at each time sample, 16 and 84 percentiles of V^* at the same time sample. The same technique was used to compute the average vergence velocity within a time window.

Since there is not a universally accepted way of computing the latency of a response, we used two different measures. The first latency measure, which we indicate as L_5 (black bars in Fig. 2), was obtained by finding the time at which DVRs to opposite sign disparity stimuli become significantly different. This is equivalent to finding the first point in time for which the 95% confidence interval (i.e., from its 2.5 to its 97.5 percentile) of the difference between the two does not straddle zero (Simon, 1995). To accomplish this robustly we actually started from the time at which

the difference was maximal, and stepped back-wards until we found the first time point at which the difference became not significant. To establish a confidence interval for the L_5 measure itself, we repeated this procedure (which as noted above involved bootstrap sampling) 1000 times, starting not from the original dataset but rather from a bootstrap sample from it. For each of these artificial datasets we computed the latency for the horizontal and vertical DVRs, so that the significance of the time difference between them could be directly assessed (paired measures).

Obviously this method might tend to overestimate the actual latency, more so when the signal-to-noise ratio is low. As an alternative, we thus fit (in the least-squares sense) the mean vergence speed V from time 0 to 10 ms after L_5 with a piece-wise linear function such that the speed estimate is zero for $t \leq L_0$ and equal to $k(t - L_0)$ for $t > L_0$ (Williams and Fender, 1977; Carl and Gellman, 1987; Krauzlis and Miles, 1996). L_0 (Fig. 2, gray bars) is then an estimate of the latency of the response. To establish a confidence interval for the L_0 measure, we repeated this procedure 1000 times, starting not from the original dataset but rather from a bootstrap sample from it. Once again, the latency of both horizontal and vertical DVRs was computed for each of these artificial datasets, allowing us to test the significance of the time difference between them. As expected, the latency estimate based on the L_0 measure was usually shorter than that based on L_5 .

Similar principles were applied to curve fitting. To compute confidence intervals on the parameters of the fit to a set of conditions, we computed a distribution of fits by selecting, 1000 times, single samples from the V^* distribution for each of the conditions belonging to the set, and fitting a curve to it. From this distribution of fit parameters we could then extract the mean and SEM for any of the parameters.

Results

DVRs to sinusoidal plaids

We measured the DVRs to a vertical sinusoidal grating (SF = 0.25 cpd, contrast 32%) to which 1° of crossed or uncrossed disparity (i.e., $\pm 90^\circ$ of phase disparity) was applied, and to a horizontal sinusoidal grating (same SF and contrast) having 1° of left- or right-hyper disparity. In Figure 3 we plot the DVRs induced by these stimuli in our three subjects. As explained in the Materials and Methods (Data analysis; Fig. 2) we actually report the difference between the DVRs to the crossed (or left-hyper) gratings and those to the uncrossed (or right-hyper) gratings. The horizontal DVR induced by the horizontal disparity of the vertical grating is plotted with black thin lines, whereas gray thin lines show the vertical DVR induced by the vertical disparity of the horizontal grating. In each case, only very small responses were induced in the orthogonal vergence channel (data not shown for clarity). As previously shown (Sheliga et al., 2006; Miura et al., 2008; Rambold and Miles, 2008), both horizontal and vertical gratings having 90° of phase disparity are effective at inducing strong DVRs, at very short latencies (in our subjects between 66 and 70 ms for vertical gratings and between 66 and 73 ms for horizontal gratings). In most subjects vertical DVRs are stronger than horizontal ones (Sheliga et al., 2006; Rambold and Miles, 2008), but the magnitude of this asymmetry is idiosyncratic (in our case larger for subjects BMS and AB than for CQ). As previously shown (Busettini et al., 2001), the dynamics of the response are also quite similar for horizontal and vertical DVRs, with the vertical DVR usually having a slightly longer latency (at most 3 ms in our subjects).

Figure 3. DVRs to type I plaids. The vergence response to the sum of two cardinal gratings (each SF = 0.25 cpd, contrast = 32%, 90° phase disparity) is quite similar to the sum of the vergence responses to the individual gratings in isolation. Thick lines indicate horizontal (black) and vertical (gray) DVRs to the plaid; thin lines indicate the horizontal DVR to a vertical grating (black) and the vertical DVR to a horizontal grating (gray). Each panel shows data from a different subject. Small line segments at the bottom of each panel indicate the latency (based on the L_5 measure) of the horizontal (black) and vertical (gray) components of the DVR to the plaid.

We then summed pairs of gratings (crossed vertical with left-hyper horizontal, and uncrossed vertical with right-hyper horizontal) to generate sinusoidal plaids. The DVR induced by the plaid (thick lines) was very similar to the sum of the DVRs to the two individual gratings. The horizontal DVR was in all subjects almost identical to the one induced by the vertical grating by itself (compare black thick and thin lines). The vertical DVR was very similar to that induced by the horizontal grating alone (compare gray thick and thin lines), being somewhat larger in one subject (BMS), almost identical in another (CQ), and somewhat smaller in subject AB. The timing of the horizontal and vertical plaid-induced DVRs was very similar, mostly reflecting the slightly different latency of the horizontal and vertical vergence channels noted above. For subject BMS the latency difference was -1 ms, for CQ and AB it was 5 ms.

This type of plaid, often referred to as a type I plaid (Ferrera and Wilson, 1990), is not particularly useful to address the questions we are interested in, since the pattern disparity in this case is simply equal to the vectorial sum of the grating disparities (measured orthogonal to the grating orientation). It is thus not possible to discern whether the system is responding to the pattern disparity, or if the horizontal (vertical) DVR simply responds to the horizontal (vertical) disparity of the vertical (horizontal) grating. To address this question we thus switched to a plaid composed of a cardinal (either horizontal or vertical) grating having $\pm 90^\circ$ phase disparity (vertical or horizontal, respectively) and an oblique grating ($\pm 45^\circ$ off the horizontal) having no disparity: a unidispersion plaid. The pattern disparity vector, being oriented along the zero disparity grating (Fig. 1D), is identical to that of the type I plaid used before. However, this particular type of plaid differs in an important way from the type I plaid we previously used: When, for example, a vertical (with horizontal disparity) and an oblique (with zero disparity) grating are summed, neither component has vertical disparity, and thus does not generate a significant vertical DVR by itself. Any vertical DVR induced by this stimulus must thus be imputed to the pattern. Furthermore, rotating the oblique grating by 90° leaves the component disparities unchanged, but reverses the sign of the vertical component of the pattern disparity. If these stimuli induce vertical DVRs, the sign of the vertical vergence component should thus reflect this sign reversal. Similarly, when a horizontal (with vertical disparity) and an oblique (zero disparity) grating are

Figure 4. DVRs to unidispairity sinusoidal plaids. When unidispairity plaids created summing a cardinal grating (with $\pm 90^\circ$ phase disparity) to a $\pm 45^\circ$ grating with zero disparity are used, the DVRs elicited are very informative. In the top row we show, for our three subjects, responses when the cardinal grating was vertical. This is graphically indicated in the legend through a stylized version of Figure 1D. The same short latency horizontal DVR was generated whether the oblique grating was oriented at $+45^\circ$ (thick black) or -45° (thin black). Ten to 15 milliseconds later a vertical DVR emerged, with opposite sign depending on the orientation of the oblique grating (thick gray: $+45^\circ$; thin gray: -45°). Neither grating in isolation can give rise to this late component, which must then be imputed to the pattern. When the cardinal grating was horizontal (bottom row), the same short latency vertical DVR was generated regardless of the orientation of the oblique grating. A delayed horizontal DVR, with a sign determined by the orientation of the oblique component, was then generated.

combined, the response induced by the pattern can be directly read off the horizontal vergence channel.

In Figure 4 we plot the responses of our three subjects to these stimuli. In the top row the plaids were composed of a vertical grating (having horizontal disparity) and a zero disparity oblique grating. The horizontal DVR to the gratings is the same whether the oblique grating has an inclination of $+45^\circ$ (thick black lines) or -45° (thin black lines). The vertical DVR has however opposite sign in the two cases (thick and thin gray lines, respectively). Two points are worth noting. First, even though neither grating has vertical disparity, there is a strong vertical DVR response, whose sign changes with the orientation of the oblique grating. As indicated above, this cannot be ascribed to either of the component disparities, and it must thus be ascribed to the pattern. Second, the vertical DVR is, in all subjects, delayed compared with the horizontal DVR. Whereas the horizontal DVR has a latency that is similar to that obtained with a single vertical grating having horizontal disparity, the vertical DVR is delayed relative to that induced by vertical disparity in a grating. The latency difference between horizontal and vertical DVR was ~ 10 – 15 ms (see Table 1 for mean and dispersion values across subjects), and always significant ($p < 0.05$).

The responses shown in Figure 4, bottom row, were instead induced by plaids composed of a horizontal grating (having vertical disparity) and a zero disparity oblique grating. In this case the vertical DVR to the gratings was the same whether the oblique grating had an inclination of $+45^\circ$ (thick gray lines) or -45°

(thin gray lines), and the latency of this response was again as short as that associated with a single grating having vertical disparity. Despite no horizontal disparity being present in either grating, a significant horizontal DVR (black lines), opposite in sign depending on the inclination of the oblique grating, was also present. Once again, this response can only be attributed to the pattern, and it was for all subjects significantly ($p < 0.05$) delayed relative to the vertical DVRs (Table 1). Since the pattern-induced DVR component is delayed regardless of whether it is horizontal or vertical, the delay observed cannot be attributed to different delays within the horizontal and vertical disparity/vergence channels (which, as noted above, are in any case significantly smaller).

Having established that the brain extracts from unidispairity plaids a disparity signal in addition to those associated with the two components, we now turn our attention to the nature of this signal. Our use of unidispairity sinusoidal plaids restricts it to two alternatives: it could result from either the disparity of the pattern itself, or the disparity of sinusoidal components that are generated when the plaid is processed through a nonlinearity (distortion products). For the moment we set aside the second possibility, and consider the first. Subsequently we will present experimental evidence that rules out any preponderant role for distortion products.

Does the brain always compute pattern disparity?

In general, pattern disparity can be extracted by performing a 2D cross-correlation between the two retinal images. With sinusoidal plaids, alternative mechanisms, such as the IOC rule, or the extraction of the disparity of monocular 2D features (“blobs”) in the stimulus, would yield the same result. Obviously, if pattern disparity were computed using such algorithms, the computation should be robust to stimulus manipulations that do not affect pattern disparity. Accordingly, decreasing or increasing (within certain limits) the spatial frequency of the oblique (zero-disparity) grating, which does not change the disparity of the pattern (or of either component), should not affect the late DVR component. We tested this hypothesis in our three subjects.

In Figure 5 we report the mean vertical vergence velocity in a 50 ms time-window starting from 80 ms (approximately the latency of the late component), as a function of the SF of the oblique grating. In this case we used a 0.5 cpd vertical grating, and varied the SF of the oblique grating by two octaves in each direction (i.e., from 0.125 cpd to 2.0 cpd). In all three subjects the magnitude of the response decreased as the frequency of the oblique grating deviated from that of the vertical grating. The SF of the oblique grating associated with the peak response, estimated from log-Gaussian fits to the data (shown in Fig. 5), was actually slightly lower than that of the vertical grating: 0.453 cpd (95% CI: 0.440–0.466) in BMS, 0.441 cpd (0.413–0.468) in CQ, and 0.48 cpd (0.428–0.533) in AB. The bandwidth (full-width at

Table 1. Delay between pattern and component disparity responses

Subject	Stimulus type	Vertical + Oblique		Horizontal + Oblique	
		L_5	L_0	L_5	L_0
BMS	Grating	12 (2.78)	11.29 (1.48)	15 (4.95)	7.33 (1.72)
	Square	11 (2.31)	9.00 (1.65)	12 (5.34)	6.73 (1.52)
	Noise	8 (2.50)	5.97 (2.31)	14 (3.04)	8.84 (1.78)
CQ	Grating	13 (5.57)	11.97 (4.41)	14 (3.56)	12.28 (2.22)
	Square	15 (2.74)	12.14 (1.71)	14 (2.43)	11.45 (2.11)
	Noise	24 (5.25)	20.80 (6.78)	22 (3.32)	17.69 (4.37)
AB	Grating	16 (3.07)	13.01 (3.81)	18 (4.99)	14.13 (3.59)
	Square	15 (4.94)	10.43 (3.64)	16 (2.15)	12.92 (2.18)
	Noise	11 (3.30)	8.45 (2.75)	20 (5.07)	13.06 (5.34)

The delay between the DVR response in the direction of the component disparity and that in the direction of the pattern disparity is shown, separately for each stimulus type, configuration, and subject. The values are expressed in milliseconds. SEM estimates for each value are also indicated (in parenthesis). Delays based on both latency measures are reported (see Statistical analysis in Materials and Methods for an explanation of the methods used). Delays based on the L_0 measure are lower than those based on the L_5 measure, mostly because of differences in the estimate of the latency of the late component, which typically had a lower signal-to-noise ratio. Each of the 36 values listed is positive and significantly different than zero ($p < 0.05$).

half-height, expressed in octaves) was 1.94 (95% CI: 1.80–2.10) in BMS, 2.14 (1.78–2.65) in CQ, and 1.96 (1.53–3.22) in AB.

With the other configuration, i.e., when a 0.5 cpd horizontal grating and an oblique grating of varying SF were paired, results were statistically the same; however, the magnitude of the horizontal DVRs was quite small even when the two gratings had similar SFs (compare Fig. 4), producing wider confidence intervals on the parameters.

To verify whether the relative SF of the two gratings is indeed the variable of interest, in two subjects we repeated the experiment reported in Figure 5, but using a range of SFs for the vertical grating, spanning a three-octave range (from 0.125 cpd to 1.0 cpd). At all frequencies of the cardinal grating tested, the pattern of responses (Fig. 6) was very similar to that reported in Figure 5, confirming that the relative SF is the main determinant of the response. However, this experiment also indicates that there is a tendency for the peak response to be shifted toward an intermediate SF. In the insets in Figure 6, we plot the oblique SF associated with the peak response (as determined from the Gaussian fits in Fig. 6) as a function of the SF of the cardinal grating. In a log-log plot they fall on a line (solid), which is slightly shallower (the slope is 0.79 in BMS and 0.77 in CQ) than the identity line (dashed). According to these fits, the maximal response is achieved for matching frequencies at 0.42 cpd for BMS and 0.35 cpd for CQ. Nevertheless, this effect of absolute spatial frequency is rather minor, and can be accounted for by the spatial frequency tuning of the disparity vergence system (Sheliga et al., 2006; Miura et al., 2008). No systematic changes in bandwidth were observed.

As noted above, this type of behavior is not what would be expected from a mechanism that computes the 2D disparity between two monocular images. It can thus be concluded that the visual system does not simply compute the pattern disparity of an arbitrary sinusoidal plaid it is presented with: a straight 2D cross-correlation between the two monocular stimuli (or any mathematically equivalent operation) is not carried out. However, a limit of this interpretation is that it does not consider the possibility that pattern disparity is computed only when the plaid is perceived as a single coherent pattern. We will next describe experimental results that rule out this explanation.

Effect of stimulus coherence

When the SF of the components of a plaid differ, pattern coherence weakens (Farell and Li, 2004), and at some point the stimu-

lus appears transparent (i.e., one grating is seen on top of the other, and the two are not perceived as being part of a single stimulus). The lack of response observed when the SF of the oblique grating is high or low (relative to the cardinal grating) could thus be reasonably imputed to the transparency of the stimulus: because no pattern is perceived, pattern disparity should not be computed. To directly test this hypothesis, which would link a perceptual phenomenon (coherence) to the DVRs, we took advantage of the observation (Farell and Li, 2004) that coherence is drastically different in plaids made up of square wave gratings (Fig. 1B). Interestingly, such plaids appear transparent when at least one of the components has some horizontal disparity, but not when only zero or vertical component disparities are present. If the SF tuning observed in our previous experiment were explained by the degree of coherence of the stimuli, we would thus expect quite different results with square wave gratings relative to those with sinusoidal gratings (Fig. 4). This did not occur: The exact same patterns of early and late DVR components observed for sinusoidal gratings was also present with square wave gratings (Fig. 7), regardless of whether the stimuli appeared transparent (top row) or not (bottom row). In fact, responses to sinusoidal and square wave gratings were even quantitatively very similar. For all conditions and subjects the latency differences between early and late components (see Table 1) were once again significantly different than zero ($p < 0.05$), and not different than those observed with sinusoidal gratings.

This demonstrates that transparency per se does not interfere with the computation of the disparity signal that drives the late DVR component: the visual system computes pattern disparity not only for coherent stimuli, but also for some transparent stimuli. Nevertheless, there are conditions under which this computation does not yield a robust pattern disparity signal (Fig. 5). The ability to use DVRs to probe the pattern disparity computation mechanism regardless of transparency, allows us, for the first time, to determine how various stimulus parameters affect the computation itself.

Two alternative mechanisms might explain the results we have presented so far. First, a 2D cross-correlation (or, equivalently, the extraction of the disparity of monocular features) might be carried out, but only after the images are passed through circular-symmetric bandpass filters (having a 2 octave bandwidth). This mechanism would operate in parallel to the orientation selective disparity energy filters (Ohzawa and Freeman, 1986; Ohzawa et al., 1990; Anzai et al., 1999) that presumably extract the component (i.e., grating) disparity responsible for the early DVR component. If this 2D mechanism were slower than the 1D one, the observed delay between early and late DVR components would be accounted for. Alternatively, component disparities might be extracted first, and would then be combined taking into account both the magnitude of the disparities and the orientation of the components. This combination would only be performed within a spatial frequency channel, and the time required to carry it out would account for the delay in the late DVR response. As we will now show, the effect of contrast on DVRs can be used to exclude any 2D cross-correlation mechanism.

Effect of stimulus contrast

Although increasing the contrast of a stimulus increases the firing rate of most visual neurons responsive to it, the sensation of depth is little affected by changes in contrast: Higher contrast stimuli appear only slightly closer or farther than low contrast stimuli (Ichihara et al., 2007; Cisarik and Harwerth, 2008; Vlaskamp et al., 2011). This presumably reflects a read-out rule

Figure 5. DVRs to unidisparsity sinusoidal plaid with different SFs. The magnitude of the pattern-related DVR component (here measured as the mean vergence speed in a time window, see Results for details) varies as a function of the frequency of the oblique grating. Here a vertical grating (SF = 0.5 cpd) with horizontal disparity was paired with a zero disparity grating (+45° orientation) whose SF varied between 0.125 cpd and 2.0 cpd. Obviously the more the SF of the oblique grating differed from the SF of the vertical grating, the more the response decreased. Filled dots are data points. Error bars indicate $\pm$ SEM. The data were fit using a log-Gaussian function with four parameters: $\text{DVR} = Ae^{-[\log(\text{SF}) - \log(\text{SF}_0)]^2 / 2\sigma^2} + b$. Each panel shows data from a different subject.

Figure 6. DVRs to unidisparsity sinusoidal plaid with different SFs: testing a range of cardinal SFs. Same as Figure 5, but in this experiment the cardinal grating could have one of four SFs: 0.125 cpd, 0.25 cpd, 0.5 cpd, or 1.0 cpd (see key). The SF of the oblique grating was varied around the SF of the cardinal grating, and best-fit log-Gaussian tuning curves like those in Figure 5 are plotted. Vertical dashed lines indicate the SF of the cardinal grating, and thus the location of the peak response expected if the SF of the two gratings were matched. As a first order approximation this holds, indicating that it is the relative SF of the two gratings that matters, not the absolute SF of the oblique grating. However, some consistent deviations can be seen. In the insets we plot (symbols) on the abscissa the SF of the cardinal grating, and on the ordinate the SF of the oblique grating that yielded the maximal DVR response. In solid we show the linear fit (in log-log coordinates) to these data, whereas in dashed we show the identity line. Only two subjects participated in this control experiment.

that discounts the impact of contrast on firing rate (Watson and Robson, 1981). Conversely, the DVRs are exquisitely sensitive to stimulus contrast, increasing linearly with its logarithm (Sheliga et al., 2006), implying a much simpler read-out rule. Indeed, it has been shown (Takemura et al., 2001) that horizontal DVRs are extremely well predicted by the difference between the summed activity of MST neurons tuned to crossed and uncrossed disparities. This sensitivity of the DVRs to contrast allows us to test a prediction of a 2D narrow-band cross-correlation mechanism: Since the 2D cross-correlation peak is determined by the contrast of the plaid (i.e., by the sum of the contrasts of the two gratings), higher plaid contrast should result in stronger responses.

To investigate this issue we manipulated independently the contrast of two gratings (0.25 cpd each) in a stereo plaid composed of a vertical sinusoidal grating having horizontal disparity, and a zero disparity oblique grating. When the cardinal and the

oblique grating had the same contrast, we found that both the horizontal (mainly component disparity driven) DVR (Fig. 8, top row) and the vertical (pattern disparity driven) DVR (Fig. 8, bottom row) increased approximately linearly with log-contrast, just as is the case if a single grating is used (Sheliga et al., 2006). Applying a different contrast to the two gratings had, however, a quite different effect on the two DVR components. When the contrast of the cardinal grating was higher than that of the oblique grating, the early DVR component (Fig. 8, top row) was mostly determined by the contrast of the cardinal grating. When the contrast of the oblique grating came to dominate that of the cardinal grating, the response was, however, reduced. Since in our experiment the cardinal grating had disparity, and thus could drive DVRs, whereas the oblique grating had zero disparity, mutual inhibition between the neural signals associated with the two gratings can readily explain this phenomenon. This would be similar to the reciprocal inhibition proposed to explain the effect of contrast on the interaction between gratings having different SFs but the same orientation (Sheliga et al., 2007; Miura et al., 2008).

The late vertical DVRs induced by pattern disparity behaved quite differently. For a given contrast of the oblique grating, the late DVR component was largest when the cardinal grating had a similar contrast (Fig. 8, bottom row): lower and, most importantly, higher contrasts of the cardinal grating (and thus of the plaid) were associated with smaller pattern disparity responses. This result rules out 2D cross-correlation as being responsible for the generation of the pattern disparity signal, and leaves an IOC-like combination mechanism between components of similar SF as the remaining candidate mechanism. Within an IOC-like framework,

mutual inhibition between the neural signals associated with the two gratings can largely explain the effect of contrast on the pattern disparity signal. It is in fact reasonable to expect that an IOC mechanism would operate optimally when the two gratings generate signals of similar strength, which under mutual inhibition requires similar contrasts.

However, a close inspection of the data reveals that even this explanation cannot fully account for our findings. Notably, under this scenario swapping the contrasts of the two gratings should have no effect on the pattern disparity computation, but this is not what we found. For example, in Figure 9 we show the time course of some representative DVRs from one subject. If we compare the responses to 5% cardinal contrast (c)–20% oblique contrast (o) vs 20%c–5%o (left), or 10%c–20%o vs 20%c–10%o (center), or 20%c–40%o vs 40%c–20%o (right), we see that in all cases both horizontal (top row) and vertical (bottom row) DVRs are different. Importantly, the early response driven by the com-

ponent disparity (top row) is stronger when the cardinal grating has higher contrast (gray lines), whereas the late response driven by the pattern disparity (bottom row) is stronger when the oblique, zero disparity, grating has a higher contrast (black lines). It is certainly reasonable to expect that the early DVR component, driven by grating disparity, would follow the contrast of the cardinal grating. It is less straightforward to explain why the late DVR component should be more sensitive to the contrast of the oblique than cardinal grating. In Discussion we will put forward a possible explanation for this finding. Of course, it would be impossible to reproduce this behavior using any 2D cross-correlation (or monocular feature disparity) scheme, since by definition it could not distinguish the contribution of the two components to plaid/feature contrast.

Role of

distortion products/nonlinearities

The results presented so far support, and significantly extend, the hypothesis put forward by Farell (1998) on the basis of depth perception experiments, that pattern disparity is computed on the basis of 1D component disparities. It has, however, been argued (Delicato and Qian, 2005) that Farell's results can also be explained by the presence of "second-order" features in sinusoidal plaid stimuli. In essence, these are contrast modulations that are not visible to a first-order mechanism, and thus could not be detected by a traditional binocular energy model (Ohzawa et al., 1990; Fleet et al., 1996). However, they become visible to such a model once the pattern is passed through a nonlinearity (such as a squaring operation). For example, in our sinusoidal plaids (Fig. 1A) it is apparent that there are oblique lines of low contrast that can be easily perceived. The argument put forward by Delicato and Qian is thus that the disparity of such components could be responsible for Farell's results, without needing to invoke a computation based on 1D first-order component disparities. Since these second-order features have considerably lower contrast than the first-order features in the image, this explanation might accommodate our observed delayed response to the pattern disparity as well. It is, however, important to note that squaring generates four second-order sinusoidal components, and to account for Farell's results only one of them must be selectively used (the one with the lowest SF).

Evaluating the impact of distortion products in psychophysical experiments is notoriously difficult, but it is considerably easier with DVRs. In fact, keeping in mind that the contrast of second-order features is always proportional to the product of the contrasts of the two gratings, the results reported in the previous section already represent a very good test. As it can be seen from Figure 8, when stimuli characterized by the same product of the two contrasts, and thus identical second-order features, are compared, the late responses (bottom row) can be considerably different (e.g., 10%–10% vs 20%–5%). These differences are

Figure 7. DVRs to unidispary square wave plaids. Same as Figure 4, but sinusoidal plaids were replaced with plaids obtained summing two square wave gratings. Top, Vertical and oblique gratings were summed, yielding a plaid that did not cohere in depth (there was clear transparency, with the vertical grating appearing in front of or behind the oblique grating; see Perception of unidispary plaids in Results). Bottom, Horizontal and oblique gratings were summed. Because the component disparity was vertical, there was no perception of transparency. Despite this major perceptual difference, the responses were very similar to those obtained with sinusoidal gratings: An early DVR in the direction of the component disparity was generated, and it was followed 10–15 ms later by a DVR with a sign that followed the orientation of the oblique component.

genuine, and are not an artifact of the method used to quantify DVRs in Figure 8 (average vergence speed in a time window). This is obvious from Figure 9, which shows the time course of the responses to plaids in which the contrasts of the two gratings were simply swapped (and thus their product was the same). No scheme based on distortion products can account for this behavior.

To further investigate this issue, we also created another set of plaids, this time composed of 1D binary noise line patterns (Fig. 1C). The line patterns making up these plaids are broadband, and thus contain a wide range of SFs. When two such line patterns, of different orientations, are summed and squared, a myriad of second-order components are introduced, smeared across the frequency and orientation spectrum (regardless of whether squaring precedes or follows orientation filtering). It is thus hard to see how the appropriate subset of such components could be selectively extracted. The strongest argument against any useful second-order signal being available from noise stimuli is, however, provided by our own visual system. If one looks at the monocular stimuli (Fig. 1), the second-order features that were so obvious with sinusoidal plaids (Fig. 1A) are nowhere to be seen with noise stimuli (Fig. 1C). One might argue that a square wave grating is also broadband, and yet second-order features are visible in square-wave plaids (Fig. 1B). However, in that case the frequency spectrum is not dense, and only the lowest SFs, which are well spaced, have a reasonably high contrast. With a square wave plaid it is thus much easier to see how a "squaring" scheme might selectively extract the second-order features of interest.

When we presented these noise stimuli to our subjects, the same patterns of early and late DVR components observed for

Figure 8. Sinusoidal plaids with different contrasts for the two components. Here a vertical grating ($SF = 0.25$ cpd) with horizontal disparity was paired with an oblique grating ($+45^\circ$ orientation same SF) with no disparity, and the contrast of the two were varied independently. The magnitude of the DVR was quantified by computing the average vertical vergence velocity in a time window (65–115 ms for horizontal vergence, top row, and 80–130 ms for vertical vergence, bottom row). Top, The component-disparity driven response increased with the contrast of the cardinal grating (abscissa). When the oblique grating had a higher contrast than the cardinal grating, the response was, however, suppressed. Bottom, When the contrast of the two gratings composing a sinusoidal unidispairity plaid was varied, the DVR elicited by the pattern disparity was maximal when the two contrasts were similar. For any given contrast of the oblique grating (different lines and symbols, see key), the maximal response was elicited when the cardinal grating (abscissa) had similar contrast: lower and higher contrasts of the cardinal grating led to reduced responses.

sinusoidal and square gratings were present (Fig. 10). All three subjects exhibited clear responses to both the component and pattern disparities. It should be noted that since our noise stimuli appear transparent (see below), this experiment reinforces our previous observation that pattern coherence is not a prerequisite for the computation of pattern disparity. Once again, there was a statistically significant difference ($p < 0.05$ in all cases) in the latency of the horizontal and vertical DVR components (Table 1). In two subjects the latency differences were very similar to those obtained with the equivalent sinusoidal and square gratings, whereas in one subject (CQ) they were ~ 8 ms longer (almost entirely imputable to an increased latency of the late response). Another difference, this time consistent across subjects, is that responses to noise plaids were always more sustained than those to sinusoidal or square wave plaids. This was true for both components of the response, but especially for the late component. Finally, relative to the early component, the magnitude of the late component was stronger with noise plaids. This was especially true when a horizontal pattern was paired with an oblique pattern (Fig. 10, bottom row). These differences presumably reflect the interaction of signals across the spatial scales represented in the noise patterns, an aspect that is worthy of future investigations.

While these experiments do not allow us to exclude that second-order features might have been responsible for part of the responses we have reported above, they certainly rule them out as being the only, or even a major, contributor to those responses.

Perception of unidispairity plaids

The results presented above are compatible with previous psychophysical investigations of pattern disparity. However, since the stimuli we used differed in some details from published psychophysical studies, we replicated certain important observations using the same stimuli that we used to elicit eye movements. One of our novel findings is that transparent patterns produce eye movements in the pattern direction. Previously it had been shown that square wave plaids in which a vertical pattern (with horizontal disparity) is paired with an oblique pattern (with zero disparity) appear transparent (Farell and Li, 2004). To confirm that some of our stimuli are also seen as transparent, we presented to four subjects (including those used for DVR measurements above) square wave and noise plaid patterns in which either the vertical or the oblique grating had disparity ($\pm 90^\circ$ phase disparity for square wave plaids, $\pm 0.2^\circ$ position disparity for noise plaids), and the other grating had zero disparity. All the stimulus parameters were identical to those used for the DVR experiment, including the stimulus duration (200 ms) and the absence of a fixation point during stimulus presentation. After the stimulus appeared, the subjects had to report (through a button press) whether the vertical grating appeared in front of or behind the oblique grating. Obviously, unless the subject perceived two gratings in different planes they could not carry

out the task. All four subjects performed the task without difficulty (85% correct or more in each condition, $p < 0.001$).

We also confirmed (in three of these subjects) that, in coherent stimuli, the pattern depth was correctly perceived. In this task, the fixation marker remained visible during the presentation of the stimulus, and subjects reported whether the pattern appeared in front of or behind the fixation marker. With sinusoidal plaids composed of a vertical (with 90° of phase disparity) and an oblique ($\pm 45^\circ$ orientation, zero disparity) grating, we found that all subjects could report ($>80\%$ correct, $p < 0.001$) whether the stimulus was near or far relative to the fixation point. When a sinusoidal horizontal grating (with 90° of phase disparity) was summed to an oblique ($\pm 45^\circ$ orientation, zero disparity) grating, only one subject (CQ) could perform the task, whereas the other two performed at chance. Lowering the vertical disparity of the horizontal grating led to improved performance, so that with 22.5° of phase disparity all three subjects perceived depth (80% correct or more, $p < 0.001$). When a square wave plaid composed of horizontal (with 90° of phase disparity) and oblique ($\pm 45^\circ$ orientation, zero disparity) gratings was used, all subjects could perform the task (80% correct or more, $p < 0.001$); lowering the disparity further improved performance. When a noise plaid composed of horizontal (with 0.2° of position disparity) and oblique (zero disparity) patterns was used, all subjects could easily perform the task (85% correct or more, $p < 0.001$), despite the absence of distortion products. Note that with any of the plaids

composed of horizontal and oblique patterns, performing the task is only possible if the plaid appears coherent, as neither component has horizontal disparity. Unlike Farell (1998), however, we have not tested conditions in which component and pattern disparity have horizontal disparities of opposite sign.

In two subjects we also tested perception under the conditions used by Farell (1998). Stimuli had higher SF (1.0 cpd), lower disparity (22.5° phase disparity, or 0.1° position disparity for the noise plaid), and were presented within a central hard-edged window (7° diameter). Presentation time was kept at 200 ms, and plaid contrast at 64%. The subjects could perform the task in all conditions tested (88% correct or more, $p < 0.001$). Notably they correctly reported the pattern disparity of the noise plaids (horizontal plus oblique patterns), revealing that even under Farell's experimental conditions distortion products are not necessary to account for the perception of pattern disparity.

Discussion

We used disparity driven vergence eye movements (DVRs) to study how the visual system extracts pattern disparity from stereo plaids. We showed that unidisparsity plaids induce robust DVRs, characterized by two components: one, at very short latency, driven by the disparity of the gratings, and another, delayed by 10–15 ms, driven by the disparity in the pattern. We thus confirmed that the visual system is capable of extracting pattern disparities from plaids, as previously reported on the basis of depth perception experiments (Farell, 1998, 2003; Chai and Farell, 2009; Farell et al., 2010). Importantly, we were able to rule out alternative interpretations, such as an effect of distortion products (Delicato and Qian, 2005). We then turned our attention to the mechanism that underlies this computation. Because with unidisparsity plaids pattern disparity cannot be obtained from any linear combination of the two component disparities, we considered two possible alternatives: a 2D cross-correlation, and a combination of component disparities based on the IOC rule (Fennema and Thompson, 1979; Adelson and Movshon, 1982). By manipulating independently the spatial frequency and contrast of the two components, we were able to rule out all 2D cross-correlation schemes. A combination of component disparities is thus the most likely explanation for our findings. We will now attempt to better characterize this mechanism in light of our results and previous findings.

A striking feature of our results is the 10–15 ms delay between the initial DVR in the direction of the component disparity and the late DVR in the pattern disparity direction. The simplest explanation for this finding is that the IOC-like neural computation of pattern disparity from component responses takes 10–15 ms. This finding parallels the observation by Masson and colleagues (Masson and Castet, 2002; Masson, 2004; Barthélemy et al., 2008, 2010) that when subjects (humans and monkeys) are presented with moving plaid stimuli the resulting Ocular Following Response (OFR) is characterized by two components: an initial one driven by the motion of the components, and a delayed one in the

Figure 9. Sinusoidal plaids with swapped contrasts between the two components. Close inspection of Figure 8 reveals that swapping the contrast of the two components in a sinusoidal plaid can lead to considerably different responses. This aspect is highlighted here by showing the time course of these responses, rather than just the average vergence speed in a time window. Top, Component driven (horizontal) DVRs. Bottom, Pattern driven (vertical) DVRs. Each column shows the responses for different pairs of contrasts (see key: %c is contrast of cardinal grating, %o is contrast of oblique grating). Note that the responses are in general significantly different, in all panels. Also note how in the top row the gray lines (higher contrast of the cardinal grating) are always larger than the black lines (higher contrast for the oblique grating). The reversed pattern is observed in the bottom row. Examples from one subject (BMS) are shown.

direction of pattern motion. In those studies, which inspired ours, the delay was of ~20 ms and thus comparable to ours. Neurophysiological evidence for such a delay has been found: pattern motion signals in area MT are delayed relative to component motion signals (Pack and Born, 2001; Smith et al., 2005, 2010). More precisely, some neurons in area MT respond to the direction of the moving gratings throughout the presentation, whereas others initially encode the component direction but subsequently respond to the pattern direction (Smith et al., 2005, 2010). Early models of pattern motion computation, which relied on selective pooling of neurons sensitive to component motion (Simoncelli and Heeger, 1998), would not produce a delay of this sort, but more recent models [such as from Rust et al. (2006)] can account for it (Smith et al., 2010). Unfortunately, there are no physiological data on pattern disparity processing. Nonetheless, in analogy with motion processing it seems reasonable to assume that binocular neurons in area V1 encode component disparity. Since DVRs appear to be mostly determined by the activity in area MST (Takemura et al., 2001, 2007), presumably at least some neurons in this area encode pattern disparity. Area MST receives large projections from area MT, where neurons are tuned for both motion direction and disparity (Poggio, 1995; DeAngelis et al., 1998; DeAngelis and Newsome, 1999). In analogy with the findings for pattern motion, it is thus tempting to speculate that neurons in MT may initially respond to component disparity, but within 15 ms a sizeable fraction of them would respond to pattern disparity instead.

Because manipulating the contrast and SF of the components of a sinusoidal plaid quickly leads to loss of coherence, psychophysical investigations of pattern disparity (or motion) can only operate within strict limits. If the computation of pattern disparity itself were restricted to coherent plaids, these limits would extend to any methodology. However, our experiments with square wave (Fig. 7) and noise plaids (Fig. 10) revealed that pat-

Figure 10. DVRs to unidispersion 1D binary noise plaids. Same as Figure 7, but here binary 1D noise line patterns were summed to obtain a plaid. Top, Vertical pattern (with $\pm 0.2^\circ$ position disparity) and oblique pattern (zero disparity) were summed. Bottom, Horizontal and oblique patterns were summed. 1D binary noise plaids are immune to the second-order artifacts that in theory might contaminate sinusoidal and square wave plaids. Nevertheless, the DVRs were qualitatively similar to those shown in Figures 4 and 7. Importantly, also in this case an early DVR in the direction of the component disparity was generated, and it was later followed by a DVR with opposite sign depending on the orientation of the oblique component.

tern disparity is computed even when the subjects perceive transparency. This implies that the computation is carried out quite early in the visual processing stream, and conversely that the sensation of transparency/coherence arises later on. Accordingly, when plaids are presented, three disparity signals (one for each component, and one for the pattern) arise. We propose that the perceptual system and the oculomotor system deal differently with the simultaneous presence of these multiple signals: whereas the vergence system is driven by the algebraic sum of these signals (Takemura et al., 2001), the perceptual system is able to access them selectively and independently. We base this proposal on three observations: First, the perceptual system must be able to disregard the component disparity signals when a coherent stereo plaid is presented, or subjects could not correctly report the depth of such plaids (Farell, 1998). Second, whereas large disparities are needed to optimally drive the disparity-vergence system, smaller disparities are more easily perceived (as shown by our perceptual tests). Finally, the perceptual system must be able to disregard the pattern disparity signals when a transparent stereo plaid is presented, or subjects could not correctly report the depth of the components (Farell and Li, 2004). It is also worth mentioning that, whereas eye movements are driven by absolute disparity, perception is driven by the comparison between two concurrently available disparity signals (relative disparity), which perforce implies an additional processing stage.

The ability to use DVRs to probe the computation of pattern disparity regardless of coherence has enabled us to study the underlying mechanism much more extensively than had previously been possible. We found that, for sinusoidal plaids, the computation of pattern disparity is very sensitive to the ratio between the

SFs of the two gratings. The strongest signal is obtained when the two gratings have similar SFs, and degrades quickly as their difference increases. From this result we infer that the IOC-like mechanism effectively combines only gratings having similar SFs (i.e., it operates across orientations but only within a spatial frequency channel).

Thanks to the exquisite sensitivity of short-latency ocular responses to contrast manipulations we were able to rule out 2D cross-correlation mechanisms (as they predicted DVRs that depend on the sum of the gratings' contrast) and distortion products (as they predicted DVRs that depend on the product of the contrasts) as being the source of the disparity signal driving the late DVR. We found instead that the most robust pattern disparity signal is obtained when the two gratings have similar contrasts. As we mentioned in Results, a mechanism of cross-orientation inhibition (Morrone et al., 1982; DeAngelis et al., 1992) might explain this outcome. This mechanism would, however, predict that when the two gratings have different contrasts, swapping them should not make a difference. This is not what we found: The late DVR is significantly stronger when the oblique grating is the one with the higher contrast in the pair. A solution to this conundrum may lie in the

observation that in our experiments the oblique grating always had zero disparity, whereas the vertical grating had $\pm 90^\circ$ of phase disparity. Under natural conditions (but not in our experimental setup), stimuli that have disparity are off the plane of fixation and thus out of focus, which would reduce their apparent contrast. To compensate for this loss of contrast, the signal associated with this component might thus be boosted; the two signals would then be of more similar strength at the input of the IOC-like computation when the signal associated with the zero disparity grating (i.e., the oblique grating in our experiments) emerged from earlier stages of processing as stronger than the vertical grating. This would imply that, under our experimental conditions, the maximal response would be obtained not for identical contrasts of the two gratings, but rather when the oblique grating has slightly larger contrast, as we observed.

In conclusion, our findings provide compelling evidence in support of the idea that pattern disparities are computed from disparities first represented in single orientation channels. This integration of information across orientation channels requires an additional 10–15 ms of processing, depends on both the relative spatial frequency and contrast of the gratings, and is carried out even when its output is not used to guide perception.

References

- Adelson EH, Movshon JA (1982) Phenomenal coherence of moving visual patterns. *Nature* 300:523–525. [CrossRef Medline](#)
- Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for processing binocular information I. Simple cells. *J Neurophysiol* 82:891–908. [Medline](#)
- Barthélemy FV, Perrinet LU, Castet E, Masson GS (2008) Dynamics of dis-

- tributed 1D and 2D motion representations for short-latency ocular following. *Vision Res* 48:501–522. [CrossRef Medline](#)
- Barthélemy FV, Fleuriot J, Masson GS (2010) Temporal dynamics of 2D motion integration for ocular following in macaque monkeys. *J Neurophysiol* 103:1275–1282. [CrossRef Medline](#)
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Busetini C, Miles FA, Krauzlis RJ (1996) Short-latency disparity vergence responses and their dependence on a prior saccadic eye movement. *J Neurophysiol* 75:1392–1410. [Medline](#)
- Busetini C, Fitzgibbon EJ, Miles FA (2001) Short-latency disparity vergence in humans. *J Neurophysiol* 85:1129–1152. [Medline](#)
- Carl JR, Gellman RS (1987) Human smooth pursuit: stimulus-dependent responses. *J Neurophysiol* 57:1446–1463. [Medline](#)
- Chai YC, Farell B (2009) From disparity to depth: how to make a grating and a plaid appear in the same depth plane. *J Vis* 9(10):3.1–19. [CrossRef Medline](#)
- Cisarik PM, Harwerth RS (2008) The effects of interocular correlation and contrast on stereoscopic depth magnitude estimation. *Optom Vis Sci* 85:164–173. [CrossRef Medline](#)
- Collewijn H, van der Mark F, Jansen TC (1975) Precise recording of human eye movements. *Vision Res* 15:447–450. [CrossRef Medline](#)
- Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24:203–238. [CrossRef Medline](#)
- DeAngelis GC, Newsome WT (1999) Organization of disparity-selective neurons in macaque area MT. *J Neurosci* 19:1398–1415. [Medline](#)
- DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. *J Neurophysiol* 68:144–163. [Medline](#)
- DeAngelis GC, Cumming BG, Newsome WT (1998) Cortical area MT and the perception of stereoscopic depth. *Nature* 394:677–680. [CrossRef Medline](#)
- Delicato LS, Qian N (2005) Is depth perception of stereo plaids predicted by intersection of constraints, vector average or second-order feature? *Vision Res* 45:75–89. [CrossRef Medline](#)
- Dobkins KR, Stoner GR, Albright TD (1998) Perceptual, oculomotor, and neural responses to moving color plaids. *Perception* 27:681–709. [CrossRef Medline](#)
- Efron B (1982) The jackknife, the bootstrap, and other resampling plans. Philadelphia: SIAM.
- Farell B (1998) Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature* 395:689–693. [CrossRef Medline](#)
- Farell B (2003) Detecting disparity in two-dimensional patterns. *Vision Res* 43:1009–1026. [CrossRef Medline](#)
- Farell B, Li S (2004) Seeing depth coherence and transparency. *J Vis* 4(3):8209–223. [CrossRef Medline](#)
- Farell B, Chai YC, Fernandez JM (2010) The horizontal disparity direction vs. the stimulus disparity direction in the perception of the depth of two-dimensional patterns. *J Vis* 10(4):25.1–15. [CrossRef Medline](#)
- Fennema CL, Thompson WB (1979) Velocity determination in scenes containing several moving images. *Comp Graph Image Proc* 9:301–315. [CrossRef](#)
- Ferrera VP, Wilson HR (1990) Perceived direction of moving two-dimensional patterns. *Vision Res* 30:273–287. [CrossRef Medline](#)
- Fleet DJ, Wagner H, Heeger DJ (1996) Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res* 36:1839–1857. [CrossRef Medline](#)
- Gorea A, Lorenceau J (1991) Directional performances with moving plaids: component-related and plaid-related processing modes coexist. *Spat Vis* 5:231–252. [CrossRef Medline](#)
- Hays AV, Richmond BJ, Optican LM (1982) A UNIX-based multiple process system for real-time data acquisition and control. Paper presented at WESCON Conference, Anaheim, September.
- Helmholtz H (1925) Treatise on physiological optics. Rochester, NY: Optical Society of America.
- Ichihara S, Kitagawa N, Akutsu H (2007) Contrast and depth perception: effects of texture contrast and area contrast. *Perception* 36:686–695. [CrossRef Medline](#)
- Julesz B (1971) Foundations of cyclopean perception. Chicago: University of Chicago.
- Krauzlis RJ, Miles FA (1996) Release of fixation for pursuit and saccades in humans: evidence for shared inputs acting on different neural substrates. *J Neurophysiol* 76:2822–2833. [Medline](#)
- Masson GS (2004) From 1D to 2D via 3D: dynamics of surface motion segmentation for ocular tracking in primates. *J Physiol Paris* 98:35–52. [CrossRef Medline](#)
- Masson GS, Castet E (2002) Parallel motion processing for the initiation of short-latency ocular following in humans. *J Neurosci* 22:5149–5163. [Medline](#)
- Michelson AA (1927) Studies in optics. Chicago: University of Chicago.
- Miura K, Sugita Y, Matsuura K, Inaba N, Kawano K, Miles FA (2008) The initial disparity vergence elicited with single and dual grating stimuli in monkeys: evidence for disparity energy sensing and nonlinear interactions. *J Neurophysiol* 100:2907–2918. [CrossRef Medline](#)
- Morrone MC, Burr DC, Maffei L (1982) Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc R Soc Lond B Biol Sci* 216:335–354. [CrossRef Medline](#)
- Ohzawa I, Freeman RD (1986) The binocular organization of simple cells in the cat's visual cortex. *J Neurophysiol* 56:221–242. [Medline](#)
- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037–1041. [CrossRef Medline](#)
- Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409:1040–1042. [CrossRef Medline](#)
- Poggio GE (1995) Mechanisms of stereopsis in monkey visual cortex. *Cereb Cortex* 5:193–204. [CrossRef Medline](#)
- Rambold HA, Miles FA (2008) Human vergence eye movements to oblique disparity stimuli: evidence for an anisotropy favoring horizontal disparities. *Vision Res* 48:2006–2019. [CrossRef Medline](#)
- Rambold HA, Sheliga BM, Miles FA (2010) Evidence from vergence eye movements that disparities defined by luminance and contrast are sensed by independent mechanisms. *J Vis* 10(14):31.1–34. [CrossRef Medline](#)
- Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE T Bio-Med Eng* 10:137–145.
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431. [CrossRef Medline](#)
- Sheliga BM, FitzGibbon EJ, Miles FA (2006) Short-latency disparity vergence eye movements: a response to disparity energy. *Vision Res* 46:3723–3740. [CrossRef Medline](#)
- Sheliga BM, FitzGibbon EJ, Miles FA (2007) Human vergence eye movements initiated by competing disparities: evidence for a winner-take-all mechanism. *Vision Res* 47:479–500. [CrossRef Medline](#)
- Simon JL (1995) Resampling: the new statistics. Alexandria, VA: Resampling Stats.
- Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vision Res* 38:743–761. [CrossRef Medline](#)
- Smith MA, Majaj NJ, Movshon JA (2005) Dynamics of motion signaling by neurons in macaque area MT. *Nat Neurosci* 8:220–228. [CrossRef Medline](#)
- Smith MA, Majaj N, Movshon JA (2010) Temporal dynamics of motion integration. In: Dynamics of visual motion processing: neuronal, behavioral, and computational approaches (Ilg UJ, Masson GS, eds), pp 55–72. New York: Springer.
- Takemura A, Inoue Y, Kawano K, Quaia C, Miles FA (2001) Single-unit activity in cortical area MST associated with disparity-vergence eye movements: evidence for population coding. *J Neurophysiol* 85:2245–2266. [Medline](#)
- Takemura A, Murata Y, Kawano K, Miles FA (2007) Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. *J Neurosci* 27:529–541. [CrossRef Medline](#)
- Vlaskamp BN, Yoon G, Banks MS (2011) Human stereopsis is not limited by the optics of the well-focused eye. *J Neurosci* 31:9814–9818. [CrossRef Medline](#)
- Watson AB, Robson JG (1981) Discrimination at threshold: labelled detectors in human vision. *Vision Res* 21:1115–1122. [CrossRef Medline](#)
- Williams RA, Fender DH (1977) The synchrony of binocular saccadic eye movements. *Vision Res* 17:303–306. [CrossRef Medline](#)