

Motion-defined Surface Segregation in Human Visual Cortex

Gabriel J. Vigano¹, Ryan T. Maloney^{1,2}, and Colin W. G. Clifford^{1,2}

Abstract

■ Surface segregation provides an efficient way to parse the visual scene for perceptual analysis. Here, we investigated the segregation of a bivectorial motion display into transparent surfaces through a psychophysical task and fMRI. We found that perceptual transparency correlated with neural activity in the early areas of the visual cortex, suggesting these areas may be involved in the segregation of motion-defined surfaces. Two oppositely rotating, uniquely colored random dot kinematograms (RDKs) were presented either sequentially or in a spatially interleaved manner, displayed at varying alternation frequencies. Participants reported the color and rotation direction pairing of the RDKs in the psychophysical task. The spatially interleaved display generated the percept of motion transparency across the range of frequencies tested, yielding ceiling task performance. At high alternation frequencies, performance on the sequential

display also approached ceiling, indicative of perceived transparency. However, transparency broke down in lower alternation frequency sequential displays, producing performance close to chance. A corresponding pattern mirroring the psychophysical data was also evident in univariate and multivariate analyses of the fMRI BOLD activity in visual cortical areas V1, V2, V3, V3AB, hV4, and V5/MT+. Using gray RDKs, we found significant presentation by frequency interactions in most areas; differences in BOLD signal between presentation types were significant only at the lower alternation frequency. Multivariate pattern classification was similarly unable to discriminate between presentation types at the higher frequency. This study provides evidence that early visual cortex may code for motion-defined surface segregation, which in turn may enable perceptual transparency. ■

INTRODUCTION

Surface segregation through the grouping of similar visual attributes is a key aspect of parsing the visual scene (Watt & Phillips, 2000; Stoner & Albright, 1996). One example of this is the perceptual decomposition of a bivectorial motion display into transparent surfaces, allowing attention to be allocated to the features of each surface (Stoner & Blanc, 2010; Valdes-Sosa, Cobo, & Pinilla, 2000). The binding of visual features is generally considered to be a relatively slow process (e.g., Treisman, 1996, 1998; Treisman & Gelade, 1980) with a perceptual limit close to 3 Hz for a stimulus alternating between two sets of features (Holcombe, 2009). However, feature binding is significantly enhanced if stimuli can be perceptually segregated into transparent surfaces (Clifford, Spehar, & Pearson, 2004; Moradi & Shimojo, 2004; Suzuki & Grabowecky, 2002; Holcombe & Cavanagh, 2001).

Moradi and Shimojo (2004) used a display that alternated between two fields of differently colored, moving dots in the same spatial location. Participants reported the color–motion conjunction present in the display. *Prima facie*, higher alternation frequencies should increase task difficulty as there is less time available to process the on-screen feature conjunction (Seymour, McDonald, &

Clifford, 2009; Bodelón, Fallah, & Reynolds, 2007). Moreover, apparent asynchronies between color and motion perception should drastically worsen performance at high alternation frequencies (Nishida & Johnston, 2002; Moutoussis & Zeki, 1997). However, Moradi and Shimojo (2004) demonstrated that an increase in alternation frequency paradoxically improved performance on the conjunction task. High alternation frequencies facilitated the perceptual grouping of the display into two motion-defined surfaces. Visual persistence is a likely mechanism for the stability of this percept, as it acts over a short temporal window (Shioiri & Cavanagh, 1992; Mezrich, 1984; Coltheart, 1980). At high alternation frequencies, motion-defined surfaces are integrated across presentations (Farrell, Pavel, & Sperling, 1990) and perceived as coherent transparent fields for the entire stimulus duration, enhancing perceptual performance.

Several studies have investigated the neurophysiological basis of motion transparency. When presented with bivectorial motion stimuli, neurons in anesthetized monkey V1 tend to respond to motion in their preferred direction, regardless of the presence of an overlapping field of dots moving in a different direction (Qian & Andersen, 1995; Snowden, Treue, Erickson, & Andersen, 1991). In contrast, the responses of V5/MT neurons are suppressed when shown the same stimuli (McDonald, Clifford, Solomon, Chen, & Solomon, 2014; Snowden et al., 1991). Human

¹University of Sydney, ²UNSW Australia

fMRI studies have demonstrated that transparent motion inhibits activity in the V5/MT+ complex in a similar manner (Garcia & Grossman, 2009; Muckli, Singer, Zanella, & Goebel, 2002; Heeger, Boynton, Demb, Seidemann, & Newsome, 1999).

Here, we investigate if the perceptual experience of transparency in bivectorial motion is matched by modulation of activity in early visual areas using a modified version of the sequential color–motion stimulus design (Moradi & Shimojo, 2004). Because changes in neural activity between stimuli presented at high and low alternation frequencies may be attributed to alternation frequency rather than the percept of transparency, we employ an additional, spatially interleaved display that appears transparent independent of frequency. Sequential and spatially interleaved stimuli are predicted to produce similar activity at higher frequencies, where they both generate the impression of motion transparency. At lower frequencies, however, the spatially interleaved stimulus should continue to appear transparent whereas the sequential stimulus will not. We therefore predict a corresponding difference in neural activity, whereby a presentation type (sequential/spatially interleaved) by alternation frequency interaction would manifest in a way that matches the perceptual experience of motion transparency.

METHODS

Participants

Informed written consent was obtained from five experienced psychophysical participants (three men; age range = 24–46 years) who participated in the psychophysical binding and fMRI components of the study. This included the three authors who were not naive to the experimental manipulations but were unaware of the condition order, in addition to two naive participants. Four participants performed an additional psychophysical subjective judgment task (three men; age range = 24–33 years), which included two of the authors, and two participants who were completely naive to the stimulus and did not participate in any other portions of this study. All participants had normal or corrected-to-normal visual acuity and normal trichromacy. Visual corrections in the MRI scanner took the form of prescription squash goggles. The experimental protocol was approved by the University of Sydney Human Research Ethics Committee.

Apparatus

Psychophysics

These experiments took place in a light- and sound-proof booth. Participants were positioned approximately 57 cm from a ViewSonic Graphics Series G90f cathode ray tube monitor (36 cm × 27 cm) with a vertical refresh rate of 60 Hz and a resolution of 1024 × 768. Thus, 1 pixel subtended 0.035° of visual angle. Stimuli were generated

through Matlab (R2010a 7.10; The Mathworks, Natick, MA) and the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) on a PC with an Intel Core i7-2600 CPU 3.40 GHz processor and an AMD Radeon HD 6450 display adapter. Color output was measured using a ColorCAL colorimeter (Cambridge Research Systems, Rochester, UK) and the LightScan software (v 2.23) and subsequently linearized using Matlab.

fMRI

Data were collected using a Philips Achieva 3T TX scanner (Philips, Amsterdam) with a whole head coil. A field-echo EPI pulse sequence was used to acquire T₂*-weighted functional MR images of BOLD contrast. The field-echo EPI protocol was defined according to the following parameters: time to echo = 32 msec, time to repetition (TR) = 3000 msec, flip angle = 90°, field of view = 192 × 69 × 192 mm, matrix = 128 × 128, voxel size = 1.5 mm (isotropic). The images were acquired in 46 interleaved ascending slices (1.5 mm thickness) in a tilted coronal plane that covered the entire occipital cortex as well as a portion of the posterior parietal and temporal lobes. In addition to the functional scans, a whole-head structural MR image (voxel size = 1 mm isotropic) was obtained for each participant within each experimental scanning session for coregistration purposes, using a turbo field echo protocol for optimal gray and white matter contrast.

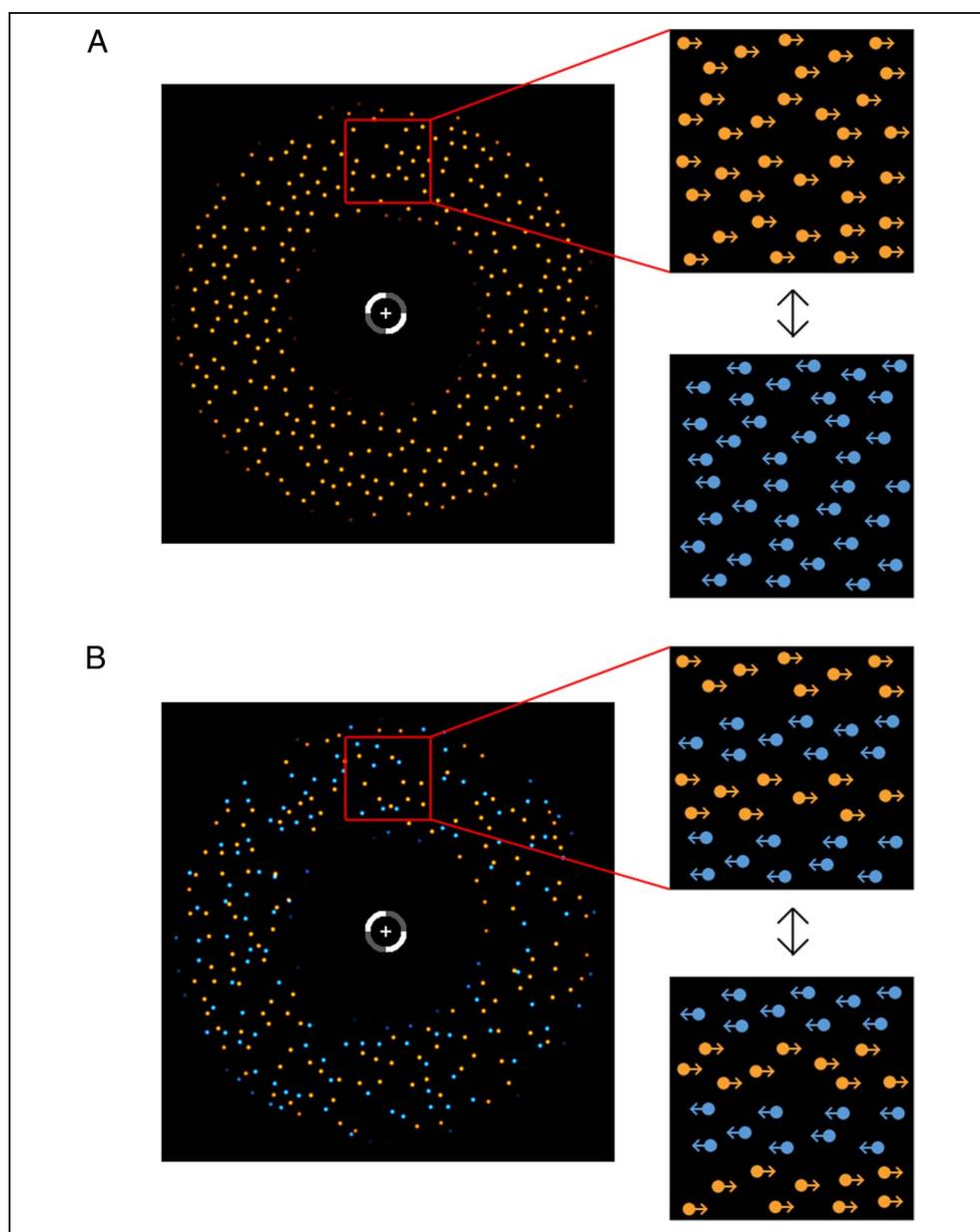
Stimuli were generated on a Dell Precision M4400 laptop with an nVidia Quadro FX 1700M display adapter and back-projected through the Faraday shield using a Dell 5100MP digital projector (Dell, Inc., Round Rock, TX) positioned behind the bore, with a resolution of 1024 × 768 pixels, a refresh rate of 60 Hz, and a mean luminance of 275 Cd/m². Images were viewed at a distance of 167 cm through a rear-facing mirror mounted upon the head coil, giving a viewing angle of 19° × 14.3° (0.019° per pixel). Scanning was performed in a darkened chamber lit only by the light from the projector. The participants' behavioral responses were collected via a LU400-PAIR Lumina response pad (Cedrus, San Pedro, CA).

Visual Stimuli

Psychophysics

Stimulus design was based on Moradi and Shimojo (2004; see Figure 1). Two random dot kinematograms (RDKs) with a luminance of 28 Cd/m², presented against a black background, were generated at the beginning of each trial: one contained orange dots (CIE: $x = 0.43, y = 0.45$) and the other blue dots ($x = 0.21, y = 0.25$). These colors were chosen using DKL color space (Derrington, Krauskopf, & Lennie, 1984) such that they summed to gray ($x = 0.29, y = 0.33$) to account for the possibility that color–motion binding performance may be biased by imbalanced color saturation or luminance. All dots had adjusted colors based

Figure 1. Stimulus design. Representation of the stimulus used in the psychophysics experiments. Both presentation types are shown: (A) the sequential display and (B) the spatially interleaved display. Arrows denote alternation. Gray dots replaced both orange and blue dots in the fMRI stimulus.



on a minimum flicker paradigm (Walsh, 1953) to ensure colors were subjectively equiluminant at approximately 28 Cd/m^2 . Dots were Gaussian blobs ($\sigma = 0.042^\circ$ of visual angle) and distributed evenly throughout the annulus with a minimum distance of 0.7° from any other dot. 11.8% of the total viewing area was filled with dots, giving a density of 3.7 dots/deg^2 . RDKs were randomly assigned opposite rotations each trial and constantly rotated at a rate of $60^\circ \text{ sec}^{-1}$ (equivalent to 0.167 Hz) while on screen. The average dot speed within the RDKs was 4.8° of visual angle per second.

The annulus containing the stimulus had a raised cosine profile with an outer radius of 6.3° of visual angle, inner radius of 2.8° and 0.8° of smoothing at both inner and outer edges. The total area of the ring was 98 deg^2 . The center of the annulus contained a gray fixation cross with

a height and length of 0.4° and line width of 0.07° . A white and gray fixation ring with a diameter of 1.4° and width of 0.5° encircled this cross. The annulus was used to minimize eye movements; rotational movement balanced motion energy across the display, in addition to eliminating transients caused by dots leaving and entering the annulus.

The psychophysical experiments had a 2 Presentation Type (sequential/spatially interleaved) \times 6 Alternation Frequency (1.67, 2.5, 3.33, 5, 7.5, and 15 Hz) within-subject design. The sequential display (Figure 1A) alternated between orange and blue RDKs such that a single color-motion combination was on screen at any single point in time. Changes in color and motion attributes occurred simultaneously. For example, the 5-Hz sequential presentation consisted of the orange RDK for 100 msec, followed

by the blue RDK for 100 msec, and so on until 1200 msec had elapsed.

The spatially interleaved display (Figure 1B) simultaneously presented both RDKs across concentric, logarithmically spaced annuli, or strips, such that a strip containing orange dots rotating in one direction was flanked on either side by strips of blue dots rotating in the opposite direction. Strip width increased linearly with radius from 0.15° to 0.39° for each of the 10 strips present in each display. Each strip alternated between orange and blue at the specified alternation frequencies. Annular strips were used to balance local motion cues and perceived flicker with the sequential presentation while avoiding dots physically overlapping. In this way, dot density was kept constant between the sequential and spatially interleaved conditions. As has been observed previously (Clifford et al., 2004) under this arrangement, separate strips from the same RDK were readily combined into a coherent whole, such that two transparent RDKs were perceived.

In all conditions, the stimulus was present for 1200 msec. A 250-msec mask was present at the beginning and end of each trial. In total, each trial had a duration of 1700 msec. The mask was a static superposition of both blue and orange dots generated in an identical fashion to the stimulus. It was used to avoid any potential transients caused by the sudden onset and offset of the stimulus, which in turn may have biased the results of the psychophysical study.

Participants completed 40 trials per condition for the binding experiment. Each condition was repeated eight times per run for a total of 96 trials per run. Five runs were completed, amounting to a total of 480 trials per participant. Within runs, each condition was counterbalanced for onset color and color–motion conjunction. Participants’ task was a color–motion “binding” task while maintaining fixation on the central cross. The rotation direction of the orange RDK (clockwise or counterclockwise) was reported on each trial using a standard keyboard.

Participants completed six runs of the subjective judgment task, for a total of 576 trials overall. Three runs used orange- and blue-colored dots, and the other three used equiluminant gray dots that were matched to those used in the fMRI study (see details below). Here, participants made two separate subjective judgments on the properties of the stimulus after the trial had ended. Participants first reported the number of surfaces they perceived simultaneously in each trial (one or two), and then they indicated the presentation type of the trial (sequential or simultaneous).

fMRI

This experiment was a 2 Presentation Type (sequential/spatially interleaved) \times 2 Alternation Frequency (5 Hz/15 Hz) within-subject design. These alternation frequencies were chosen for the fMRI experiment as they had optimal temporal characteristics for maximizing the perceptual dif-

ference between presentation types, as determined through the psychophysical portion of the study (see below). Sequential conditions were compared with frequency-matched spatially interleaved displays. This was to ensure the fMRI results obtained for the sequential presentations were not an artifact of frequency differences and in fact due to differences in surface perception.

Receptive field sizes in monkey visual cortex increase as a function of both eccentricity and processing hierarchy (Smith, Singh, Williams, & Greenlee, 2001; Gattass, Gross, & Sandell, 1981; Zeki, 1978). Smith et al. (2001) combine estimates of receptive fields of primary visual areas from previous neurophysiological studies. From this, the approximate receptive field sizes of several visual areas within the eccentricity of the stimulus annulus can be estimated. Receptive fields in V1 at an eccentricity of $2.8\text{--}6.3^\circ$ are approximately 0.5° of visual angle, V2 is 1.5° , V3 is 2.5° , and V4 is 4.5° . These data agree with estimates of receptive field size in humans using fMRI (Dumoulin & Wandell, 2008; Smith et al., 2001).

Color has previously been shown to be an effective surface segregation cue (Perry & Fallah, 2012; Croner & Albright, 1997; Edwards & Badcock, 1996). As this was an investigation into motion-defined surfaces, other features that could potentially demarcate the two RDKs may confound the fMRI results. Gray RDKs were therefore used to ensure the observed results were purely because of motion-defined surfaces. The results of the subjective judgment task (reported in the next section) indicate that there were no significant differences in the way participants perceived motion transparency between the colored stimuli used in the primary psychophysical experiment and grayscale stimuli used in the fMRI experiment. In addition to this, no mask was used between stimulus presentations. Apart from these aspects, stimuli were identical to the psychophysical experiment (refer to Figure 1).

Stimuli were presented in 15-sec counterbalanced blocks, wherein five 3-sec timed repetitions (TRs) took place. In each block, five repetitions of stimuli (2500 msec) and fixation (500 msec) were presented. This was done to keep presentation times comparable to the psychophysical task. Here, stimuli had a 500-msec raised cosine contrast ramp on and off instead of a mask. As there was no color–motion conjunction present in this stimulus, a stimulus mask was not needed. Block order was counterbalanced both between and within runs. Stimulus blocks were presented in groups of four, with 15-sec fixation blocks between them. Each condition was presented four times per run, for a total of 21 blocks per run. Runs lasted for 315 sec, and participants viewed 12 runs in total.

To control for attention and fixation, participants performed an attentionally demanding dimming task throughout each run. In the center of the display, the fixation cross alternated between black and gray on average every 1500 msec, jittered randomly by ± 500 msec. Participants indicated (by holding down one of the two buttons on

the response pad) the current state of the fixation cross (dimmed or not dimmed). Task duration, maintenance of fixation, target size, and frequency of contrast changes were all contributing factors to ensure the task was not trivial, requiring fixation. Button press data were used to quantify participants' ability to maintain fixation during their time in the scanner.

Analysis

Retinotopic Mapping/Definition of ROIs

In previous sessions, both functional and high-resolution anatomical scans were acquired from each participant. An average anatomical image was prepared consisting of whole-head sagittal and transverse images (voxel size = 1 mm isotropic) and a higher-resolution partial coronal image (voxel size = 0.75 mm isotropic) of the caudal brain to maximize anatomical detail in the occipital lobes. Before averaging, the images were aligned using normalized mutual information-based coregistration, inhomogeneity corrected (Manjon et al., 2007), and normalized according to their peak white matter intensities and resampled (where necessary) to a voxel size of 0.75 mm (isotropic). Each participant's average anatomical image was then segmented using the automatic algorithms of ITK-SNAP (Yushkevich et al., 2006; www.itksnap.org/) and mrGray (Teo, Sapiro, & Wandell, 1997), supplemented with careful manual editing.

Functional scans were obtained of participants viewing clockwise/counter clockwise rotating wedges and expanding/contracting ring stimuli as described in Wandell, Dumoulin, and Brewer (2007). Data were coregistered through SPM5 (www.fil.ion.ucl.ac.uk/spm/software/spm5/; Friston, Ashburner, Kiebel, Nichols, & Penny, 2007) and organized into ROIs. The maximal activations of each voxel to the wedge stimuli was then used to generate a polar angle map of the visual cortex using the best-fitting sinusoid for the time course of each voxel (for more details, see Larsson & Heeger, 2006). From this map, visual areas were manually defined in mrVista (white.stanford.edu/software). Functionally defined early visual cortex was delineated for each participant using the nomenclature and criteria of Wandell et al. (2007) and Larsson and Heeger (2006), in the same manner as previous studies from this laboratory (see Supplementary Figure 1 in Mannion, McDonald, & Clifford, 2010). According to this scheme, areas V1-V3 and hV4 share a foveal representation at the occipital pole, whereas V3A and V3B (which were not separated in this analysis) share a dorsal foveal representation and border the dorsal portion of V3. Area hV4 was defined as a hemifield representation of the contralateral visual field bordering the ventral portion of V3 (Goddard, Mannion, McDonald, Solomon, & Clifford, 2011). In separate localizer scans, area V5/MT+ was localized as a region of lateral visual cortex in the ascending limb of the inferior temporal sulcus responding to coherently moving versus static random dot stimuli presented at low contrast (Dumoulin et al., 2000).

Data Analysis

Preprocessing of the functional data acquired for this study was done using the same methods as the retinotopic analysis above. SPM5 was used to average, coregister, and reslice the functional scans to the same space as the ROIs. Using a general linear model (GLM) contrast of fixation versus stimulus blocks, we generated a maximum intensity projection of the most informative voxels to be used as a mask for the ROIs in this analysis. These were the voxels that had the greatest change in response averaged over all stimulus blocks compared with fixation. Note that this contrast is orthogonal to those between the conditions of interest.

The univariate analysis was conducted by calculating BOLD signal time courses for each voxel per condition, shifted forward by two TRs (6 sec) to account for the response lag because of the hemodynamic response. Aguirre, Zarahn, and D'Esposito (1998) and others report an average peak hemodynamic RT approximating 5 sec. We shifted our data the integer number of TRs that most closely matched this measurement. The mean signal time course for each condition was calculated by averaging all voxels in each ROI across all runs. The average signal of the fixation blocks was subtracted from this result to express the result as a function of signal change from fixation.

The multivariate analysis was similar in that responses were again grouped by condition, averaged over runs, and shifted by two TRs. For each voxel, the time series of responses to each stimulus block within each run was z-scored (fixation blocks were not used in the analysis), and a response to each block was computed as the mean of the z scores from the five corresponding TRs within a single block. A linear support vector machine (SVM) as implemented in SVMlight (Joachims, 1999; C parameter set to 1.0) compared the difference in the pattern of activation for each visual area between the sequential and spatially interleaved conditions of a single alternation period. Eleven runs were used as training data, and the twelfth was used as a test. For each visual area, this process was repeated 12 times, such that all permutations of test and training assignments were run. The performance of the SVM was recorded for each visual area for each participant as the percentage of correct classifications of test blocks.

Button-press data for the fMRI experiment were assessed using a mutual information (MI) analysis, given by

$$MI = \sum_{r,s} p(s)p(r|s) \log_2 \frac{p(r|s)}{p(r)}$$

where $p(r)$ is the probability of a response, $p(r|s)$ is the probability of a response given the stimulus, and MI is expressed in bits of information. Response data were time-shifted with respect to the time course of the fixation cross luminance to find the lag (representing RT)

generating the maximal amount of mutual information. Mutual information was then converted to efficiency, given by

$$\text{Efficiency} = \frac{100 \times \text{MI}}{H(s)}$$

where $H(s)$ is the entropy of the stimulus display train, defined as

$$H(s) = \sum_s p(s) \log_2 p(s)$$

Efficiency denotes the degree to which the stimulus predicts the response, with a higher efficiency indicating a better correspondence between stimulus and RT courses, reflecting appropriate fixation. The mean response lag across participants was 430 msec. Measured at a resolution of 60 Hz, the minimum efficiency in a single run averaged across participants was 48.9% ($SD = 16.2\%$), and the maximum was 74.4% ($SD = 3.3\%$). Minimum efficiencies were well above chance, suggesting participants were fixating in all runs of the study. Mean efficiency averaged across all runs ranged from 46.9% to 69.6% between subjects. If participants were responding at random, an efficiency close to 0% would be expected.

RESULTS

Psychophysics

The psychophysical experiments investigated the effect of presentation type and alternation frequency on surface segregation. This was quantified through the color-motion binding task (Figure 2) and corroborated with subjective reports of stimulus interpretation in a separate experiment (Figure 3). The spatially interleaved stimulus was designed to appear transparent regardless of alterna-

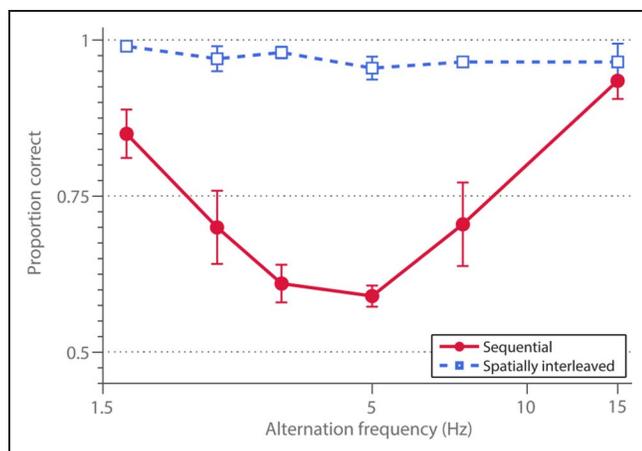


Figure 2. Color-motion binding task results. Performance on the color-motion binding task averaged over participants ($n = 5$) for each condition. Error bars denote ± 1 between-subject standard errors.

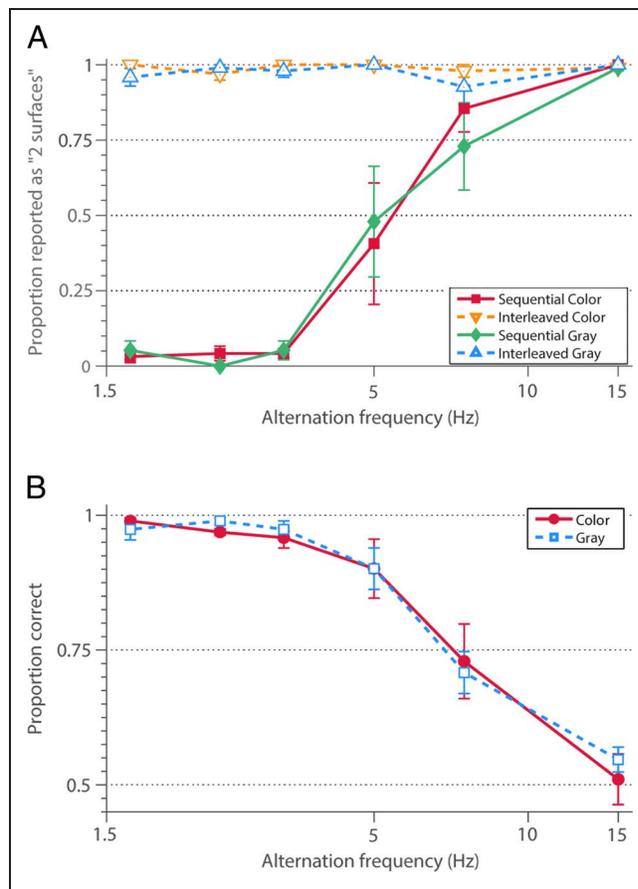


Figure 3. Subjective judgment task results. Performance on the subjective judgment tasks averaged over participants ($n = 4$) for each condition. Both the surface judgment (A) and the display identification judgment (B) tasks are displayed here. Error bars denote ± 1 between-subject standard error.

tion period, which was supported by the lack of a linear trend averaged across spatially interleaved conditions in Figure 3A, $F(1, 3) = 0.01$, $p = .93$. Matched binding performance between the display conditions and the subjective judgment task suggested an equivalence in perception between the sequential and spatially interleaved displays at higher temporal frequencies. Evidence for this is provided in Figure 3B, with identification of displays decreasing linearly as alternation frequency increased, $F(1, 3) = 168.80$, $p = .001$, indicating that sequential and spatially interleaved conditions tended to be less distinguishable at higher alternation frequencies. Therefore, task difficulty was directly correlated to the perception of multiple motion-defined surfaces: higher performance was associated with motion-defined surface segregation.

Figure 2 displays binding task performance. Using a repeated-measures GLM, we found a main effect of both Presentation Type, $F(1, 4) = 171.07$, $p < .001$, and Alternation Period, $F(1, 4) = 8.11$, $p = .001$. Critically, we also found a significant quadratic Presentation \times Frequency interaction, $F(1, 4) = 84.60$, $p = .001$. Although performance in the spatially interleaved conditions was at

ceiling, performance in the sequential display decreased and then increased across frequency (Moradi & Shimojo, 2004). As both presentation types had the same dot density, local motion cues, and alternation period, the difference in performance can therefore be attributed to an interaction between presentation type and frequency. A clear trend was present in the sequential display data, $F(1, 4) = 107.97, p < .001$. Performance was close to ceiling in the sequential display at the highest (15 Hz) and lowest (1.67 Hz) alternation frequencies and approached chance as frequency approached 5 Hz. The spatially interleaved display generated ceiling performance for all alternation frequencies, correlating with the reported perception of two transparent motion-defined surfaces.

This is further emphasized by the difference in subjective impressions between sequential and spatially interleaved displays in Figure 3A. The significant effect of Alternation Frequency on condition recognition, $F(1, 3) = 55.51, p < .001$, is correlated with the number of surfaces reported in Figure 3A. That is, as alternation frequency increased, participants were more likely to report that they distinguished two surfaces, when averaged over display type, $F(1, 3) = 235.47, p = .001$. Furthermore, there was a significant Display Type \times Frequency interaction effect, averaged across both colored and gray displays, $F(1, 3) = 229.46, p = .001$, such that only the sequential display produced a varying percept as a function of alternation frequency. The presence of both motions in the interleaved display generates the impression of two continuously present surfaces, which then enabled the color–motion pairings to be isolated and identified in the binding task. This also occurred for the sequential display, but exclusively at high frequencies where it also produced the percept of two simultaneous surfaces.

There were no observed differences or interactions observed between Color and Gray displays in either of the subjective judgment tasks, suggesting the psychophysical and fMRI stimuli resulted in very similar perceptions of transparent motion. No significant differences between Color and Gray displays were observed in the sequential display, $F(1, 3) = 0.44, p = .55$, or the interleaved display, $F(1, 3) = 1.85, p = .27$, and furthermore, no significant differences in the identification of display types were found, $F(1, 3) = 0.15, p = .72$.

fMRI

The aim of the fMRI experiment was to probe the neural substrates correlated with motion-defined surface perception. All conditions except the 5-Hz sequential condition were perceived as two surfaces rotating in opposite directions, as evidenced by the psychophysical data, so we might expect to see a presentation by frequency interaction. This was indeed what was found in the univariate analysis (Figure 4A). A repeated-measures GLM (uncorrected) was used to assess the data for each visual area. Significant interaction effects in visual areas V1 ($F(1, 4) = 46.59, p = .002$),

V2 ($F(1, 4) = 34.27, p = .004$), V3 ($F(1, 4) = 19.10, p = .012$), V3AB ($F(1, 4) = 14.31, p = .019$), and hV4 ($F(1, 4) = 9.55, p = .037$) but not V5/MT+ ($F(1, 4) = 0.05, p = .84$) were found. Univariate activity in early visual cortex was found to vary differently across frequencies for presentation type. This specific Presentation \times Frequency interaction suggests that, similar to the psychophysical data, activity in early visual areas modulates with respect to the conscious perception of motion-defined surface segregation.

We examined the effects present in the univariate results in further detail using multivariate pattern analysis (Figure 4B), specifically to determine whether the 15-Hz sequential and spatially interleaved displays could be identified based on the elicited patterns of activity in each visual area. An SVM classifier maximizes the chances of detecting an effect, such as differences in BOLD signal patterns or activation, which may not be evident in the univariate analysis. One-sample *t* tests were used to assess the performance of the SVM. At the 15-Hz alternation period, the SVM was only able to classify conditions significantly above chance in area V3AB, $F(4) = 34.521, p = .004$. This is in contrast to the 5-Hz alternation period, where a one-sample uncorrected *t* test demonstrated that the SVM was able to classify conditions significantly above chance using patterns of activity from nearly all areas (V1: $F(4) = 55.77, p = .002$; V2: $F(4) = 95.85, p = .001$; V3: $F(4) = 13.97, p = .020$; V3AB: $F(4) = 28.97, p = .006$; V5/MT+: $F(4) = 18.93, p = .012$) except hV4, $F(4) = 7.62, p = .051$. This suggests that not only the overall level of activation but also the patterns of activation in early visual areas mirror our perception of motion-defined surfaces.

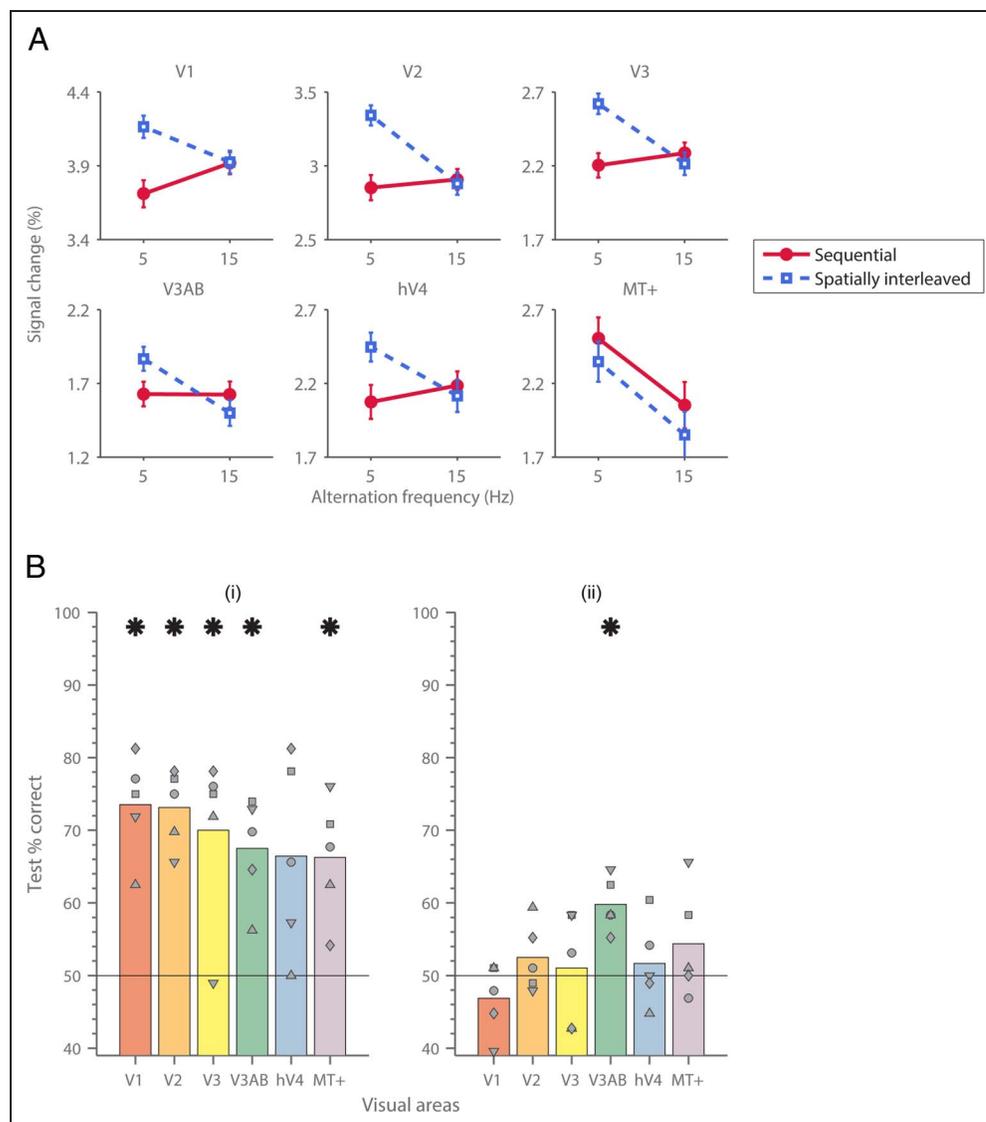
Comparisons between sequential and spatially interleaved display at 15 Hz in both univariate and multivariate analyses found little differences in activity generated. In contrast, there were large differences between presentation types at 5 Hz in both analyses, where the spatially interleaved display still looked transparent whereas the spatially interleaved display did not. These results support the notion that activity in early visual cortex correlates with the perception of motion-defined surfaces.

DISCUSSION

This study investigated the temporal dynamics of motion-defined surface segregation using psychophysics and fMRI. In both experiments, there was a clear interaction between the presentation type (sequential and spatially interleaved) and the alternation frequency. Psychophysical color–motion binding task performance was at ceiling across all frequencies in the spatially interleaved presentation type. Performance in the sequential presentation type matched that of the interleaved display at 15 Hz, coinciding with the maximal value on the surface judgment task where participants perceived displays as most transparent.

The segregation of the two RDKs into transparent surfaces in both the sequential and spatially interleaved conditions may involve motion opponency at the early stages of

Figure 4. Univariate and multivariate fMRI results. (A) Average change in univariate BOLD responses of early visual areas across all participants ($n = 5$) relative to fixation blocks. Note that each plot has identical y-axis scales, but different ranges. Error bars indicate between-subject standard errors. (B) Average performance of an SVM classifier distinguishing between sequential and spatially interleaved displays at (i) 5 Hz and (ii) 15 Hz. Bars indicate average performance in decoding conditions. Small gray icons indicate classifier performance for individual participants. Asterisks denote significantly greater classification (all p s < .05) than chance (50% indicated by the black horizontal line) across participants.



visual processing (Jones, Grieve, Wang, & Sillito, 2001; Lindsey & Todd, 1998) and/or an imbalance of local motion cues (Qian, Andersen, & Adelson, 1994). At this high frequency, visual persistence may also contribute to the perceived transparency of the sequential condition. Despite the serial presentation of RDKs, if presentation intervals occur within a sufficiently short temporal window (Shioiri & Cavanagh, 1992), both RDKs may appear to persist simultaneously, producing the percept of simultaneous, transparent, and distinct surfaces. At alternation frequencies around 5 Hz, individual color–motion presentations were too short to identify and bind features together, and performance was near chance (see also Moradi & Shimojo, 2004). At these frequencies, the RDKs may mask each other. Unlike the high-frequency conditions, the stimulus duration here exceeded the time course of visible persistence, preventing the perception of motion transparency (Shioiri & Cavanagh, 1992; Coltheart, 1980).

Binding performance steadily improved with a reduction in the alternation frequency below 5 Hz. Although only one

surface was now perceived at a time, a longer alternation period enabled features to be bound within a single presentation (Moradi & Shimojo, 2004; Nishida & Johnston, 2002). Feature binding is believed to be a relatively slow process (Treisman & Gelade, 1980). Therefore, without surface segregation, a long presentation period is required to reliably perceive feature pairings (Holcombe, 2009; Treisman, 1996, 1998; Treisman & Gelade, 1980). Together, these results suggest that either a high alternation frequency or simultaneous presentation, as in the case of the spatially interleaved condition, is necessary to enable stable transparent surface segregation (Moradi & Shimojo, 2004).

Behavioral data indicated that the display conditions appeared more similar with an increase in alternation frequency, although the physical differences between them remained identical. We found this trend in the univariate analysis of the fMRI data, revealing neural correlates of surface transparency in visual cortical areas as early as V1. Significant interaction effects of a form that mirrored

the psychophysical results were observed in all areas except V5/MT+. Activity between sequential and interleaved 15-Hz conditions was not significantly different, whereas at 5-Hz large differences in overall activity were observed. Neurons in monkey primary visual cortex respond to the presence of their preferred direction independent of transparency (Qian & Andersen, 1995; Snowden et al., 1991). Qian et al. (1994) propose that motion transparency is due to unbalanced local motion cues at the spatial resolution of V1. However, this alone would not account for the presentation type by frequency interaction effect found in this study. The distribution of local motion signals within the stimuli was identical in both presentation types across time. Given the temporal integration properties and small receptive field sizes of neurons in V1, a difference in activity at a local level between presentation types is not necessarily expected (Snowden et al., 1991). As motion transparency necessitates an integration of motion signals across the display, the differences in frequency-dependent activation between conditions and motion transparency more likely arise because of interactions with areas containing larger receptive fields (Stoner & Blanc, 2010; Dubner & Zeki, 1971). In this way, the modulation of activity in early visual areas may be reflective of not just low-level spatio-temporal filtering (Gegenfurtner, Kiper, & Fenstemaker, 1996; Leventhal, Thompson, Liu, Zhou, & Ault, 1995; Foster, Gaska, Nagler, & Pollen, 1985) but also feedback from higher-level areas (Bouvier & Treisman, 2010; Stoner & Blanc, 2010; Sajda & Finkel, 1995), although the present results do not allow us to test this speculation.

It is important to consider both the spatial and temporal properties of visual neurons (Gegenfurtner et al., 1996; Leventhal et al., 1995; Foster et al., 1985), as it is their interaction that generates the perception of motion transparency within the specific range of parameters measured here. The higher amplitude response to the spatially interleaved presentation type may be a result of the greater motion contrast within this condition (Heeger et al., 1999; Shulman, Schwarz, Miezin, & Petersen, 1998; Tynan & Sekuler, 1984), as some cells will have receptive fields on the borders of the concentric motion strips. These cells would be expected to contribute differently to the population response sampled with fMRI compared with their response to the spatially uniform RDKs in the sequential display, especially at lower alternation frequencies. Some neurons in macaque V1, for instance, show significant center-surround organization that bears a strong comparison with similar mechanisms in V5/MT and may have a role in extracting motion contrast information (Jones et al., 2001). Regardless of the precise underlying mechanism, the differences in the fMRI responses represent the result of the spatiotemporal filtering of the stimuli by populations of visual cortical neurons, and here, we show this is correlated with the perception of transparent motion as measured behaviorally. In other words, by altering only the temporal properties of our stimuli, we were able to affect the formation of motion-

defined surfaces, and this corresponded with activity in V1 and subsequent areas.

Multivariate pattern analysis further highlighted the distinction between displays at 5 Hz and 15 Hz. Classifier performance at 5 Hz produced significantly above chance decoding of presentation type from the patterns of activity in most areas, with hV4 approaching but not reaching significance. However, high decoding performance at 5 Hz is not surprising here, given the large differences in the intensity of the response between presentation types apparent in the univariate analysis. The majority of visual areas we examined appear to respond similarly to sequential and spatially interleaved displays at 15 Hz despite the physical differences between them. Multivariate pattern analysis failed to discriminate between conditions at 15 Hz for most visual areas, indicating that similar activity between conditions in the univariate analysis was not due to different patterns of activation simply averaging out to the same overall level of activity.

Previous studies have found that the perception of motion transparency modulates activity in V5/MT+ (Garcia & Grossman, 2009; Muckli et al., 2002; Treue, Hol, & Rauber, 2000; Heeger et al., 1999). Here we found a consistent multivariate result, but no significant univariate interaction. Overall activity in V5/MT+ averaged over presentation type decreased from 5 to 15 Hz, which is consistent with the operation of dynamic inhibitory processes. A combination of large receptive fields (Albright & Desimone, 1987; Dubner & Zeki, 1971) and mutual inhibition by opposing directions of motion (Stoner & Blanc, 2010; Qian et al., 1994) could account for this result, as pooling of local motion signals sampled from V1 would result in little net motion. Despite a lack of univariate modulation, patterns of activity enabled differential decoding of presentation type at 5 Hz but not 15 Hz in V5/MT+, highlighting the value of performing the more sensitive multivariate analyses in addition to the conventional univariate ones. Together with the univariate result, this suggests that V5/MT+ may play a role in coding surface segregation independent of overall levels of activity.

V3AB was the only area to produce above chance decoding at 15 Hz. This may indicate that V3AB is receptive to the physical motion of the stimulus at timescales of less than 70 msec. V3A and V3B have both been found to modulate strongly when motions from individual objects are perceptually grouped together in a transparent fashion (Caplovitz & Tse, 2010). V3B also includes portions of the kinetic occipital region, which responds more strongly to spatially segregated than transparent motion (VanOostende, Sunaert, VanHecke, Marchal, & Orban, 1997). This evidence is consistent with the suggestion that V3AB is sensitive to the fine temporal structure in the 15-Hz displays used here.

In a previous study of visual feature binding, Seymour, Clifford, Logothetis, and Bartels (2009) were able to decode color-motion conjunctions in human visual cortex as early as area V1 using a multivariate pattern analysis similar to the one used here. They used a transparent motion

stimulus in which oppositely moving sets of dots were presented simultaneously with different colors. Presenting stimuli in this way with different color–motion pairings ensured that displays could not be decoded based on the presence or imbalance of any particular visual feature. Here, we found that activity in early visual cortex was correlated with the perception of motion-defined surface transparency. This supports the notion that the results of Seymour, Clifford, et al. (2009) reflect the decoding of representations of differently colored surfaces, rather than local color–motion conjunctions.

In summary, activity in early visual areas correlated with the percept of motion-defined surface transparency. Early visual areas are implicated in the perception of motion transparency, which in turn plays a major role in visual feature binding. Together, these results are consistent with the idea that bound features are coded as surfaces in V1 and subsequent visual areas.

Acknowledgments

This work was supported by an Australian Research Council (ARC) Future Fellowship (C.W.G.C.; grant FT110100150), a National Health and Medical Research Council grant (C.W.G.C.; Grant APP1027258), and the ARC Centre of Excellence in Vision Science. The authors thank Dr. Tamara Watson for help with data analysis and Kirsten Moffatt and the MRI radiography team at St. Vincent's Hospital, Darlinghurst.

Reprint requests should be sent to Gabriel J. Vigano, School of Psychology (A19), University of Sydney, Sydney, New South Wales, Australia, 2006, or via e-mail: gvig3764@uni.sydney.edu.au.

REFERENCES

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage*, *8*, 360–369.
- Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research*, *65*, 582–592.
- Bodelón, C., Fallah, M., & Reynolds, J. H. (2007). Temporal resolution for the perception of features and conjunctions. *Journal of Neuroscience*, *27*, 725–730.
- Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science*, *21*, 200–204.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Caplovitz, G. P., & Tse, P. U. (2010). Extrastriate cortical activity reflects segmentation of motion into independent sources. *Neuropsychologia*, *48*, 2699–2708.
- Clifford, C. W. G., Spehar, B., & Pearson, J. (2004). Motion transparency promotes synchronous perceptual binding. *Vision Research*, *44*, 3073–3080.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*, 183–228.
- Croner, L. J., & Albright, T. D. (1997). Image segmentation enhances discrimination of motion in visual noise. *Vision Research*, *37*, 1415–1427.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate-nucleus of macaque. *Journal of Physiology*, *357*, 241–265.
- Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, *35*, 528–532.
- Dumoulin, S. O., Bittar, R. G., Kabani, N. J., Baker, C. L., Le Goualher, G., Pike, G. B., et al. (2000). A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcal patterning. *Cerebral Cortex*, *10*, 454–463.
- Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*, 647–660.
- Edwards, M., & Badcock, D. R. (1996). Global-motion perception: Interaction of chromatic and luminance signals. *Vision Research*, *36*, 2423–2431.
- Farrell, J. E., Pavel, M., & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research*, *30*, 921–936.
- Foster, K., Gaska, J. P., Nagler, M., & Pollen, D. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *The Journal of Physiology*, *365*, 331–363.
- Friston, K. J., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (2007). *Statistical parametric mapping*. London: Academic Press.
- Garcia, J. O., & Grossman, E. D. (2009). Motion opponency and transparency in the human middle temporal area. *European Journal of Neuroscience*, *30*, 1172–1182.
- Gattass, R., Gross, C., & Sandell, J. (1981). Visual topography of V2 in the macaque. *Journal of Comparative Neurology*, *201*, 519–539.
- Gegenfurtner, K. R., Kiper, D. C., & Fenstemaker, S. B. (1996). Processing of color, form, and motion in macaque area V2. *Visual Neuroscience*, *13*, 161–172.
- Goddard, E., Mannion, D. J., McDonald, J. S., Solomon, S. G., & Clifford, C. W. G. (2011). Color responsiveness argues against a dorsal component of human V4. *Journal of Vision*, *11*, 1–21.
- Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., & Newsome, W. T. (1999). Motion opponency in visual cortex. *Journal of Neuroscience*, *19*, 7162–7174.
- Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, *13*, 216–221.
- Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*, 127–128.
- Joachims, T. (1999). *Making large-scale SVM learning practical*. Cambridge, MA: MIT Press.
- Jones, H., Grieve, K., Wang, W., & Sillito, A. (2001). Surround suppression in primate V1. *Journal of Neurophysiology*, *86*, 2011–2028.
- Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience*, *26*, 13128–13142.
- Leventhal, A., Thompson, K., Liu, D., Zhou, Y., & Ault, S. (1995). Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *The Journal of Neuroscience*, *15*, 1808–1818.
- Lindsey, D. T., & Todd, J. T. (1998). Opponent motion interactions in the perception of transparent motion. *Perception and Psychophysics*, *60*, 558–574.
- Manjon, J. V., Lull, J. J., Carbonell-Caballero, J., Garcia-Marti, G., Marti-Bonmati, L., & Robles, M. (2007). A nonparametric MRI inhomogeneity correction method. *Medical Image Analysis*, *11*, 336–345.
- Mannion, D. J., McDonald, J. S., & Clifford, C. W. G. (2010). Orientation anisotropies in human visual cortex. *Journal of Neurophysiology*, *103*, 3465–3471.

- McDonald, J. S., Clifford, C. W., Solomon, S. S., Chen, S. C., & Solomon, S. G. (2014). Integration and segregation of multiple motion signals by neurons in area MT of primate. *Journal of Neurophysiology*, *111*, 369–378.
- Mezrich, J. J. (1984). The duration of visual persistence. *Vision Research*, *24*, 631–632.
- Moradi, F., & Shimojo, S. (2004). Perceptual-binding and persistent surface segregation. *Vision Research*, *44*, 2885–2899.
- Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society, Series B, Biological Sciences*, *264*, 393–399.
- Muckli, L., Singer, W., Zanella, F. E., & Goebel, R. (2002). Integration of multiple motion vectors over space: An fMRI study of transparent motion perception. *Neuroimage*, *16*, 843–856.
- Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology*, *12*, 359–368.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Perry, C. J., & Fallah, M. (2012). Color improves speed of processing but not perception in a motion illusion. *Frontiers in Psychology*, *3*, 92.
- Qian, N., & Andersen, R. A. (1995). V1 responses to transparent and nontransparent motions. *Experimental Brain Research*, *103*, 41–50.
- Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. 1. Psychophysics. *Journal of Neuroscience*, *14*, 7357–7366.
- Sajda, P., & Finkel, L. H. (1995). Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience*, *7*, 267–291.
- Seymour, K., Clifford, C. W. G., Logothetis, N. K., & Bartels, A. (2009). The coding of color, motion, and their conjunction in the human visual cortex. *Current Biology*, *19*, 177–183.
- Seymour, K. J., McDonald, J. S., & Clifford, C. W. G. (2009). Failure of colour and contrast polarity identification at threshold for detection of motion and global form. *Vision Research*, *49*, 1592–1598.
- Shioiri, S., & Cavanagh, P. (1992). Visual persistence of figures defined by relative motion. *Vision Research*, *32*, 943–951.
- Shulman, G. L., Schwarz, J., Miezin, F. M., & Petersen, S. E. (1998). Effect of motion contrast on human cortical responses to moving stimuli. *Journal of Neurophysiology*, *79*, 2794–2803.
- Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*, 1182–1190.
- Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, *11*, 2768–2785.
- Stoner, G. R., & Albright, T. D. (1996). The interpretation of visual motion: Evidence for surface segmentation mechanisms. *Vision Research*, *36*, 1291–1310.
- Stoner, G. R., & Blanc, G. (2010). Exploring the mechanisms underlying surface-based stimulus selection. *Vision Research*, *50*, 229–241.
- Suzuki, S., & Grabowecky, M. (2002). Overlapping features can be parsed on the basis of rapid temporal cues that produce stable emergent percepts. *Vision Research*, *42*, 2669–2692.
- Teo, P. C., Sapiro, G., & Wandell, B. A. (1997). Creating connected representations of cortical gray matter for functional MRI visualization. *IEEE Transactions on Medical Imaging*, *16*, 852–863.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *353*, 1295–1306.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Treue, S., Hol, K., & Rauber, H.-J. (2000). Seeing multiple directions of motion—physiology and psychophysics. *Nature Neuroscience*, *3*, 270–276.
- Tynan, P., & Sekuler, R. (1984). Simultaneous motion contrast: Velocity, sensitivity and depth response. *Vision Research*, *15*, 1231–1238.
- Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 488–505.
- VanOostende, S., Sunaert, S., VanHecke, P., Marchal, G., & Orban, G. A. (1997). The kinetic occipital (KO) region in man: An fMRI study. *Cerebral Cortex*, *7*, 690–701.
- Walsh, J. W. T. (1953). *Photometry* (2nd ed.). London: Constable.
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*, 366–383.
- Watt, R. J., & Phillips, W. A. (2000). The function of dynamic grouping in vision. *Trends in Cognitive Sciences*, *4*, 447–454.
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, *31*, 1116–1128.
- Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, *277*, 273–290.