Neuropsychologia 48 (2010) 2699–2708

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Extrastriate cortical activity reflects segmentation of motion into independent sources

Gideon P. Caplovitz[a,*], Peter U. Tse[b]

[a] Princeton Neuroscience Institute, 3-N-13 Green Hall, Princeton University, Princeton, NJ 08544, United States
[b] Department of Psychological and Brain Sciences, HB 6207 Moore Hall, Dartmouth College, Hanover, NH 03755, United States

ARTICLE INFO

Article history:
Received 14 January 2010
Received in revised form 2 April 2010
Accepted 9 May 2010
Available online 15 May 2010

Keywords:
Rotational motion
Form motion interaction
V3v
V4v
V3A
V3B
Retinotopic cortex

ABSTRACT

Identical local image motion signals can arise from countless object motions in the world. In order to resolve this ambiguity, the visual system must somehow integrate motion signals arising from different locations along an object's contour. Difficulties arise, however, because image contours can derive from multiple objects and from occlusion. Thus, correctly integrating respective objects' motion signals presupposes the specification of what counts as an object. Depending on how this form analysis problem is solved, dramatically different object motion percepts can be constructed from the same set of local image motions. Here we apply fMRI to investigate the mechanisms underlying the segmentation and integration of motion signals that are critical to motion perception in general. We hold the number of image objects constant, but vary whether these objects are perceived to move independently or not. We find that BOLD signal in V3v, V4v, V3A, V3B and MT varies with the number of distinct sources of motion information in the visual scene. These data support the hypothesis that these areas integrate form and motion information in order to segment motion into independent sources (i.e. objects) thereby overcoming ambiguities that arise at the earliest stages of motion processing.

1. Introduction

The ability to encode and accurately represent motion is one of the fundamental functions of the visual system. Determining how this is accomplished is a primary goal of visual research. The problems that must be resolved in order to construct an accurate motion percept are non-trivial and arise at the earliest stages of visual processing. Due to the receptive field properties of neurons early in the visual processing stream, the local detection of motion is intrinsically ambiguous. An infinite number of possible real-world motions can produce identical local responses in neural populations that detect motion. The problem of solving this many-to-one mapping is commonly referred to as the 'aperture problem' (Adelson & Movshon, 1982; Fennema & Thompson, 1979; Marr, 1982; Nakayama & Silverman, 1988a, 1988b). It has been widely hypothesized that these ambiguities are resolved by processes that integrate local motion signals in the image together to produce percepts that, for the most part, accurately reflect the actual motion of an object in the visual scene (e.g. Adelson & Movshon, 1982; Bonnet, 1981; Burt & Sperling, 1981; Hildreth, 1984; Watson & Ahumada, 1985; Weiss & Adelson, 1998, 2000; Weiss, Simoncelli, & Adelson, 2002). Models such as these are based on the integration of motion signals arising along the contours of individually moving objects.

However, there is a problem with such models. It is often the case that the visual scene contains many moving objects, portions of whose contours may not be visible due to lighting, occlusion, or other environmental factors. As such, in order to properly integrate locally ambiguous motion signals, the visual system must first resolve which signals arise from the motion of which object. This process of segmentation defines the parameters within which mechanisms of motion integration may operate. Depending on the solution to the object segmentation problem, the manner in which motion signals are integrated together can lead to dramatically different perceptual outcomes that may or may not accurately reflect what is actually occurring in the world.

The motion of a rotating ellipse provides a simple perceptual framework for investigating these processes of form segmentation and motion integration. The spatio-temporal sequence of retinal images produced by a rotating ellipse is simultaneously consistent with rigid rotational and non-rigid deformational motion (Vallortigara, Bressan, & Bertamini, 1988; Wallach, Weisz, & Adams, 1956; Weiss & Adelson, 2000). However, despite this intrinsic ambiguity, observers generally perceive only one of these percepts for a given ellipse. Specifically, a high aspect-ratio 'skinny' ellipse rotating about its center in the 2D plane will most likely be perceived to rotate rigidly, whereas a low aspect-ratio 'fat' ellipse will most

* Corresponding author. Tel.: +1 215 495 5503; fax: +1 603 646 1419.
E-mail addresses: gcaplovi@princeton.edu (G.P. Caplovitz), Peter.Tse@dartmouth.edu (P.U. Tse).

0028-3932/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2010.05.017

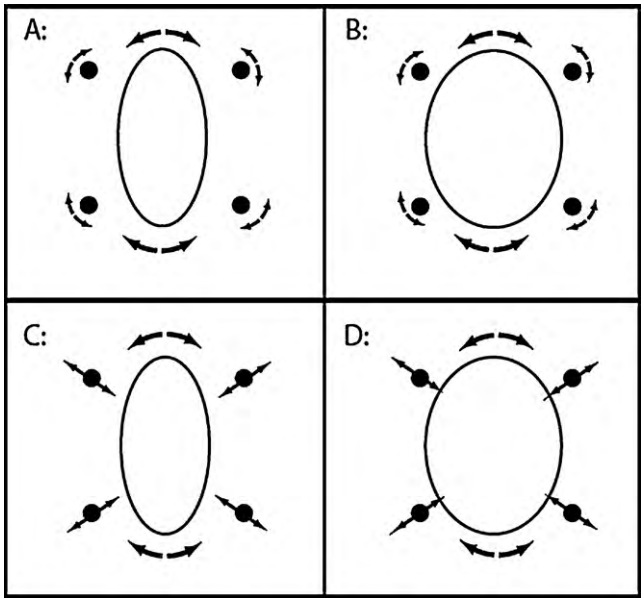

Fig. 1. Satellites and ellipses. When satellites rotate along with either a (A) skinny ellipse or (B) a fat ellipse, the percept of the ellipse's motion is that of rigid rotation. However, if the satellites translate in a radial manner, only the skinny ellipse (C) will appear to rotate rigidly. The fat ellipse (D) will be perceived to non-rigidly deform.

likely be perceived to deform non-rigidly, as if its contour were elastic, giving it a gelatinous appearance (Vallortigara et al., 1988; Wallach et al., 1956; Weiss & Adelson, 2000).

It has been hypothesized that this perceptual dissociation between rigid and non-rigid percepts reflects the mechanisms by which local motion signals are integrated along the elliptical contour (Grzywacz & Yuille, 1991; Hildreth, 1984; Horn & Schunck, 1981; Weiss & Adelson, 2000). For example, it has been hypothesized that locally detected motion signals are integrated across non-local regions in the image such that the resultant output will be optimized relative to certain constraints such as smoothness (Grzywacz & Yuille, 1991; Hildreth, 1984; Horn & Schunck, 1981; Weiss & Adelson, 2000). According to these models, local motion signals can be integrated such that the resultant global motion for a skinny ellipse is more consistent with rigid rotation, and the global motion for a fat ellipse is more consistent with non-rigidity or deformation.

Are the rigid and gelatinous cases distinct perceptual outcomes that are mediated by common neural mechanisms, as proposed by such models? Or are these two percepts mediated by distinct neural processes, contrary to such models? This is the first of two questions we seek to answer with the present research. Employing fMRI, we directly compare BOLD signals in response to stimuli that produce rigid and non-rigid responses. If these percepts are mediated by distinct neural mechanisms then we would expect to identify specific regions of visual cortex in which the responses differentiate between these two possible perceptual outcomes for the same stimulus.

Intriguingly, it has been shown that the motion of objects (satellites) distal to the elliptical contour can influence whether or not rigid rotation is perceived (Weiss & Adelson, 2000). In particular, a 'fat' ellipse can be made to appear to rotate rigidly by adding satellites that rotate with the same angular velocity, as if the dots were attached to the ellipse via invisible rods. If the satellites instead translate in a radial manner, maintaining a constant distance from the elliptical contour, as shown in Fig. 1, the same fat ellipse will appear to deform non-rigidly (Weiss & Adelson, 2000). In contrast, if the same satellite motion trajectories are added to a display in which a skinny ellipse is rotating, no such influence is observed. That is, no matter what the satellite trajectory is, the skinny ellipse will always be perceived to rotate rigidly. An example of the effects of satellites on ellipse rigidity can be observed in supplemental video 1.

This dissociation reflects the segmentation problem that the visual system must resolve in order to distinguish the motion of one object from another. Understanding the nature of these segmentation and integration processes and how they relate to percepts of rigid and non-rigid motion and motion perception in general has been the subject of both psychophysical research (e.g. Burr, Baldassi, Morrone, & Verghese, 2009; Caplovitz & Tse, 2006; Kohler, Caplovitz and Tse, 2009; Lorenceau & Shiffrar, 1992; Tse & Logothetis, 2002; Tse, 2006; Verghese & Stone, 1996; Weiss & Adelson, 2000) and computational modeling (Berzhanskaya, Mingola and Grossberg, 2007; Weiss & Adelson, 2000). What are the neural mechanisms that underlie these segmentation and integration processes? This is the second question we seek to address with the current research.

Existing neuroimaging research into the neuronal basis of motion segmentation has largely focused on the motion signals arising from translating dot fields (Muckli, Singer, Zanella, & Goebel, 2002) or drifting gratings (Castelo-Branco et al., 2002) that can either be integrated into a single coherently translating motion field or segmented into multiple transparent layers. Castelo-Branco et al. (2002) found increased activity in hMT+ and left V3/V3A when superimposed drifting gratings were segmented and perceived as transparent layers compared to when they were integrated into a coherently moving plaid. Similarly, Muckli et al. (2002) found increased activity in hMT+, the right posterior intraparietal sulcus and a portion of the fusiform gyrus when translating dot fields were segmented into transparent layers. Are the same mechanisms as those underlying the segmentation of translational motion signals into independent layers recruited for the case of moving objects and the segmentation of rotational motion into independent layers? The distinct challenges that rotational motion poses to the visual system suggest that additional processes may need to be recruited in order to resolve the ambiguities of rigid and non-rigid motion.

Here we employ fMRI to localize the mechanisms underlying the motion segmentation processes that mediate the influences of satellites on the disambiguation of the rigid and non-rigid percepts that can arise from a single continuously rotating ellipse. Specifically, we contrast the difference in BOLD signal activation between (1) conditions in which radial and rotational satellites move in conjunction with skinny ellipses, where radial satellite motion does not induce ellipse non-rigidity, with (2) conditions in which the same satellites move in conjunction with fat ellipses, where radial satellite motion does induce ellipse non-rigidity.

Hypotheses: Because the motions of the satellites and elliptical contours are only segmented into separate sources of motion in the radial skinny ellipse condition, we hypothesize that areas of visual cortex that are selectively involved in the processes associated with segmentation of motion into independent sources will show differential activity in the two skinny ellipse conditions (Fig. 2 top) and not in the two fat ellipse conditions. On the other hand, since the non-rigid percept is only present in the radial fat ellipse condition we hypothesize that areas of visual cortex selectively involved in dissociating rigid and non-rigid motion will show differential activity in the two fat ellipse conditions and not in the two skinny ellipse conditions (Fig. 2 bottom).

Summary of results: We find differential activity selective for the skinny ellipse conditions in several areas of extrastriate cortex. This activation likely reflects mechanisms underlying processes associated with the segmentation of image motions into independent sources or objects. In contrast, our analyses yield no evidence for regions of visual cortex that selectively dissociate rigid from non-rigid motion percepts.

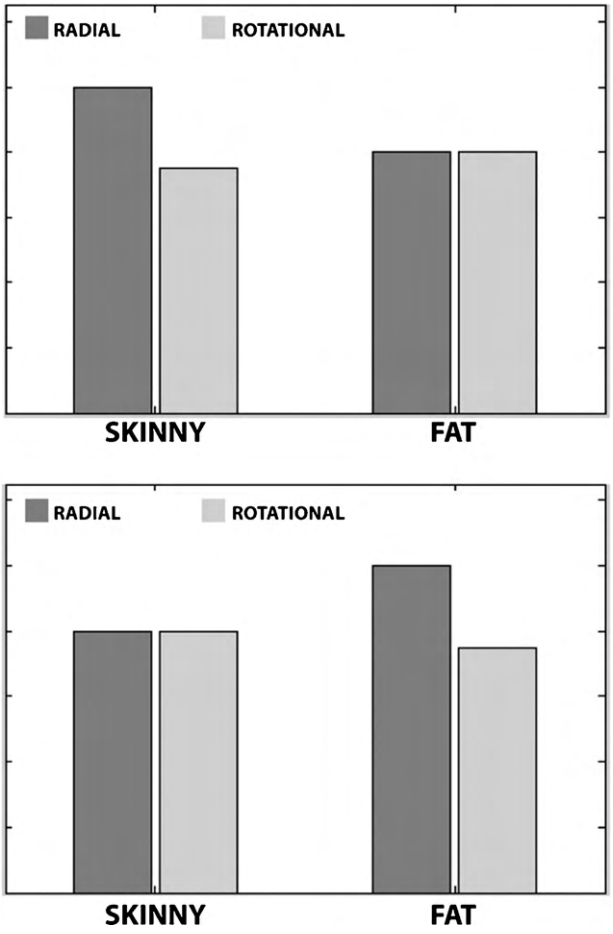

Fig. 2. Predictions. Top: Predicted BOLD signal responses for an area that is selectively involved in the processes associated with segmenting the motion signals arising from more than one object. Here differential (whether greater or lesser) activity is predicted only for the skinny ellipses with increased activity in the radial condition arbitrarily chosen to reflect increased processing necessary to segment and represent the motions of multiple objects. Bottom: Predicted BOLD signal responses for an area that is selectively involved in the dissociation and representation of rigid and non-rigid motion. Here differential activity (whether greater or lesser) is predicted only for the fat ellipses with increased activity in the radial condition arbitrarily chosen to reflect increased processing necessary to represent non-rigid motion.

2. Methods

2.1. Participants

Eleven (five male, six female, seven Caucasian, and four Asian) right handed with normal or corrected to normal vision observers participated in this experiment. Nine of these observers returned to participate in the secondary control experiment. Each observer gave informed written consent prior to participating in the experiment, and was paid $20 per scanning session.

2.2. Stimulus presentation

Stimuli were generated using the STIM software package running in Windows 2000 and presented in the scanner using the Visuastim (Resonance Technologies) binocular goggle system with a 60-Hz refresh rate.

2.3. Scanning parameters

In each of the localizers and the experiments described below, continuous whole brain BOLD signal was acquired at the Dartmouth Brain Imaging Center on a Philips 3T Achieva Intera scanner using an eight-channel SENSE head coil. Functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (3.0 mm × 3.0 mm in-plane voxel resolution, matrix size = 80 × 80, 3.5 mm thickness, axially oriented slices, interleaved slice acquisition, 0.5 mm gap, FOV = 240 × 240, TE = 35 ms, flip angle = 90°; sense factor of 2). A minimum of 8 s of dummy scans were collected at the beginning of every scan in order to allow transient activity resulting from the initiation of the scans to return to baseline. A T1-weighted coplanar anatomical image with the same slice orientation as the EPI data was also collected. This image was used to aid in the manual co-registration of functional data to a high-resolution anatomical scan that was collected in a separate scanning session. This high-resolution scan was the same as that used to construct the flattened cortical meshes used to define the regions of interest.

2.4. fMRI preprocessing

Functional MRI data were analyzed using BRAIN VOYAGER QX V1.86. Effects of small head movements during and between runs were removed using BV's motion correction algorithms. Slice scan-time correction was carried out to correct for the fact that slices were not collected at the same time and were collected in interleaved and ascending order. Functional data were not smoothed in the space domain, but any low-frequency temporal fluctuations whose wavelength was greater than 1/3 the duration of the run were removed. This did not introduce correlations between a voxel and its neighbors. For each participant, the functional data were manually co-registered to the same high-resolution anatomical image that was used to construct the cortical meshes used in defining the regions of interest.

2.5. Fixation task

A central fixation square was present during every run of each experiment. This central fixation square briefly changed color ~30 times per run. Observers were instructed to maintain central fixation at all times and respond via a separate button press every time the fixation color changed. We computed the percentage of fixation changes that were responded to within 3 s of the change on each run. Runs in which the percentage of correct responses fell below 75% were rejected from further analysis.

2.6. Defining regions of interest

A region of interest approach was used to analyze all fMRI data collected in the experiments. Regions of interest were defined using localizer scans collected in separate recording sessions from those used to collect the experimental data. Each localizer scanning session was used to identify specific functionally defined regions of visual cortex for each participant independently. Regions of interest were defined for each of the 11 observers who participated in the main experiment. The sections that follow describe the procedures by which these regions were identified and defined.

2.6.1. Creation of flattened cortical meshes

For each participant, high-resolution anatomical images were acquired using a high-resolution 3D magnetization-prepared rapid gradient echo sequence (MPRAGE; 160 sagittal slices, 0.938 mm × 0.938 mm in-plane voxel resolution, matrix size = 256 × 256, slice-thickness = 1.0 mm, FOV = 240 × 240 × 160 m; TE = 4.6 ms, TR = 9.9 ms, flip angle = 8°). The data were interpolated to provide a voxel size of 1 mm × 1 mm × 1 mm. For segmentation purposes, the scans were re-oriented to symmetrically align with the anterior commissure (AC)–posterior commissure (PC) plane. Because the individually defined region of interest approach does not require co-alignment across participants, no further spatial normalization was performed. For each participant, the orientation and geometry of their AC–PC aligned anatomical scan served as the common space for all the functional data (for both the localizers and main experiment) acquired for that participant.

In order to clearly identify each of the functionally defined regions of interest, functional data from the localizers were projected onto flattened cortical meshes that were created on the basis of the anatomical scan. The first stage of creating the flattened cortical meshes is the identification of the white matter/grey matter boundary. After separating the two hemispheres, an automatic segmentation algorithm based on signal intensity implemented in Brain Voyager QX was used to allow for an initial identification of the boundary independently for each hemisphere. Due to subtle inhomogeneities in white and grey matter intensity, in some regions the algorithm would fail to identify clearly visible boundaries. On a voxel-by-voxel basis, manual corrections were made so that the derived boundary was located slightly within the grey matter on a consistent basis across visual cortex for each hemisphere of each participant. This corrective procedure allows for boundaries that more accurately reflect the underlying topography of the cortex.

The corrected boundary was then used to create a cortical mesh from the voxels identified as belonging to grey matter. This mesh was spatially smoothed and inflated using automatic algorithms in Brain Voyager QX. The occipital portion of the mesh was cut, re-aligned and projected into the 2D plane. An automatic linear distortion correction algorithm was then applied to correct for distortions arising from the 3D to 2D projection. The degree of linear distortion due to the projection is influenced in part by the way in which the inflated mesh was cut. An attempt to minimize the residual linear distortion was made by cutting and re-cutting the meshes until the residual linear distortion was no greater than 18 (minimum possible = 13) as reported by Brain Voyager QX. This process was repeated for both hemispheres of each participant.

2.6.2. Retinotopic mapping

Retinotopy was carried out on each observer who participated in the study using standard phase-encoding techniques with the modification that two wedges of an 8 Hz flickering black and white polar checkerboard grating were bilaterally opposite like a bowtie, to enhance signal to noise (Sereno et al., 1995; Slotnick & Yantis, 2003). Functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (32 axially oriented slices, TR = 2000 ms).

Wedges occupied a given location for one TR (2.0 s) before moving to the adjacent location in either a clockwise or counter-clockwise fashion (direction was alternated across runs). Each wedge subtended 22.5° of 360°. Six TRs of dummy scans were discarded before each run to bring spins to baseline. 168 volumes were collected on each run. Five runs of each direction were collected for each participant and then averaged to minimize noise before retinotopic data analysis in BV. In addition, three runs were collected per participant using expanding 8 Hz flickering concentric rings that each spanned approximately 1° of visual angle in ring width. For every TR (2.0 s) a given ring was replaced by its outward neighbor, except the outermost ring that was replaced by the innermost ring, whereupon the cycle was repeated. These rings were used to locate the central foveal convergence area, as well as the distinct common foveal representation for V3A and V3B. Retinotopic areas (V1, V2d, V2v, V3d, V3v, V4v, and V3A/B) were defined as masks on the basis of standard criteria (Sereno et al., 1995), assuming a contralateral quadrant representation for V2d, V2v, V3d, and V3v, and a contralateral hemifield representation for V1, V4v, and V3A/B (Tootell et al., 1997). The hemifield representation just anterior to V4v, called VO1 (Brewer, Liu, Wade, & Wandell, 2005), was likely combined with what we called V4v because the border between these regions was not distinct in most of the subjects. In 9 of the 11 participants a distinct V3A/B fovea could be identified, allowing separate masks to be created for V3A and V3B. Supplemental Figure 1 illustrates an example of how the activations from the wedge and ring runs are used to perform the retinotopic identifications.

2.6.3. Localizing hMT+, MT and MST

Individual hMT+ masks and sub-area MT and MST masks (Supplemental Figure 2) were made following procedures outlined by Huk, Dougherty, and Heeger (2002). The masks were constructed by analyzing data collected across three scanning sessions: one each for hMT+, MT and MST. In each session, functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (27 axially oriented slices, TR = 1500 ms).

hMT+ masks were identified by contrasting BOLD signal activity generated by stimulus blocks (6 TRs in which 64 randomly positioned expanding and contracting white dots, each with a radius of 0.45°) with stationary blocks (18 TRs in which the 64 dots were stationary). The direction of motion (expansion/contraction) in the motion condition changed every 1.5 s. Each run consisted of 96 TRs, with an additional 6 dummy TRs collected at the beginning of each run. In each run, four blocks of each condition were presented in an interleaved fashion with the run always beginning with a motion block and ending with a stationary block. Three runs total were collected from each participant. hMT+ was identified in the left and right hemispheres as an isolated cluster of contiguous voxels in which greater BOLD signal was observed in the motion condition as compared to the stationary condition at a threshold of at least $p < 10^{-20}$ uncorrected and localized to the junction of the inferior occipital gyrus and inferior temporal sulcus.

MST masks were identified using a very similar procedure as for hMT+ except that the dots were always positioned so that the center of expansion/contraction was 7.5° of visual angle to one side of the center of the display while fixation was maintained 7.5° to the other side of the display. The overall size of the array of dots never exceeded 15° in radius, and as such, all stimulation was presented to one visual hemifield at a time. Six total runs were collected, three runs in which the dots were positioned to the right of fixation interleaved with three runs in which the dots were presented to the left of fixation. MST masks were identified for the left and right hemispheres as isolated clusters of contiguous voxels in which greater ipsilateral BOLD signal was observed in the motion condition as compared to the stationary condition at a threshold of $p < 10^{-8}$ (uncorrected) and located either overlapping with hMT+ or just superior and lateral to it. Thus the left hemisphere MST was identified using the stimuli presented to the ipsilateral left visual field and the right hemisphere MST was identified using the stimuli presented to the ipsilateral right visual field.

Like hMT+, MT was localized using 64 randomly positioned dots centered in the middle of the display surrounding the centrally located fixation point. During each run, the dots located within a 90° wedge expanded and contracted for 3 s, changing direction after 1.5 s. Every 3 s, the wedge in which the dots moved rotated by 20°. This process repeated itself until the wedge had completed six complete 360° cycles. Three runs were collected for each participant, then averaged to minimize noise before further analysis. A linear correlation analysis similar to the phase-encoding technique used in standard retinotopic mapping was performed to identify those voxels for which the moving dots within a given area of the visual field produced greater activity than for the stationary dots. MT was identified as an isolated cluster of contiguous voxels that showed contralateral and retinotopic organization either overlapping with hMT+ or located just medial to it.

Distinct activations corresponding to MT and MST were identified for 9 of the 11 participants.

2.6.4. Localizing the kinetic occipital (KO) area

Functional MRI studies have identified a cortical region designated as the kinetic occipital (KO) areas located between retinotopic area V3B and hMT+. KO has been shown to be particularly responsive to motion-defined or kinetic borders (Van Oostende, Sunaert, Van Hecke, Marchal, & Orban, 1997).

Individual KO masks were made following standard procedures (Van Oostende et al., 1997). Functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (32 axially oriented slices, TR = 2000 ms). The stimulus was a standard kinetic contour localizer consisting of motion-defined contours (Van Oostende et al., 1997). A null condition consisted of a full field of random luminance-defined noise moving uniformly back and forth. The kinetic and uniform conditions were alternately presented in blocks of 20 s. Over the course of a run, four blocks of each condition were presented. Observers were presented with a minimum of 8 runs during a recording session. Runs alternated with either the kinetic or uniform conditions being presented first. The masks in each case were determined by contrasting the kinetic > uniform conditions using a statistical threshold of at least $p < 0.0001$ uncorrected, constrained to the flattened cortical meshes of the corresponding participant. KO was identified as a clearly identifiable cluster located between hMT+ and V3B. Distinct activations corresponding to KO were identified in 8 of the 11 participants.

2.6.5. Localizing the lateral occipital complex (LOC)

Individual LOC masks were made following standard procedures (Kourtzi & Kanwisher, 2000a). Functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (32 axially oriented slices, TR = 2000 ms). Object images ($7° \times 7°$) were placed on a white background, and were embedded within a black grid. Control images were comprised of the same images scrambled within the same grid. Their centroid position was updated randomly every TR within a 1° radius of the fixation point in order to prevent perceptual fading. A central fixation spot was present at all times. Stimuli were presented in 20-s blocks separated by blank periods during which only the fixation spot was present. In a given run there were four stimulus blocks: two with unscrambled images and two with scrambled images. Three runs were collected per participant. Functional data from a given participant were co-registered to the high-resolution anatomical scan used to create the flattened cortical meshes for that participant. Left and right hemisphere LOC masks were created from the fixed-effects GLM analysis contrast of unscrambled objects > scrambled objects for each participant. The masks in each case were determined using a statistical threshold of at least $p < 0.05$ uncorrected constrained to the flattened cortical meshes of the corresponding participant.

In some participants there is an anterior-ventral portion of the LOC located in the middle fusiform gyrus in addition to a more posterior and dorsal portion located just inferior to hMT+ that is activated by this contrast. The present LOC masks were selected as the posterior region, since the two subregions were not abutting in any participant, and could well comprise areas with different functionalities. Distinct activations corresponding to the LOC were identified in 9 of the 11 participants.

2.6.6. Removal of overlapping voxels

Due to the fact that the boundaries between individually localized areas are manually defined one at a time, voxels that lie along a given boundary can mistakenly be assigned to both of the areas sharing the boundary. Due to a lack of an objective approach for assigning any such voxels to one or the other boundary-sharing areas, these voxels were identified and removed from the masks defining both areas. For both hemispheres overlapping voxels were removed from the following (potential) boundaries: V1–V2d, V1–V2v, V2d–V3d, V2v–V3v, V3d–V3A, V3d–V3B, V3A–V3B, V3v–V4v, V4v–LOC, V3d–KO, V3B–KO, KO–LOC, KO–MT, KO–MST, MST–MT, LOC–MT, LOC–MST. While this led to the discarding of voxels that were localized in retinotopic and other localizers, which lowered statistical power, this step is conservative in that it guarantees that BOLD signal responses measured in a given region of interest arise only from this one area.

2.7. Experimental stimuli

The stimuli for the main experiment were presented continuously in 20-s blocks. Each stimulus consisted of four identical rotating ellipses and four corresponding sets of four moving dot-satellites. As shown in Fig. 1, ellipses were either skinny (high aspect-ratio) or fat (low aspect-ratio), and the satellites traversed either rotational or radial trajectories. Each ellipse was positioned inside one of the cardinal visual quadrants. The center of each ellipse was located $6° \times 4.5°$ visual angle ($H \times V$) away from the center of the display. The ellipses and dots were white and presented against a grey background. Each of the dots had a radius of 0.15° of visual angle. In every condition the ellipses rotated back and forth around their center, oscillating between −20° and 20° from vertical at an angular velocity of $75° \, s^{-1}$). In any given block, the aspect-ratio of the rotating ellipses was either high (subtending $3.6° \times 1.8°$ visual angle) or low (subtending $3.6° \times 2.7°$) and the motion of the satellites was either rotational (at the same angular velocity as the ellipse) or radial (such that a constant distance between each dot and the contour of the fat

ellipse was maintained). It is important to note that the radial trajectories were defined on the basis of the fat ellipse even for the skinny ellipse condition. Thus, as demonstrated in the supplemental video, both the rotational and radial motion of the satellites was identical for the high and low aspect-ratio conditions. When presented with the rotational-motion satellites, both the high and low aspect-ratio ellipses were perceived to rotate rigidly. When presented with the radial motion satellites, the fat ellipses were perceived to deform. In contrast, with the radial motion satellites, the skinny ellipses were perceived to rigidly rotate. The perceived rigidity of the skinny ellipse in the radial condition is not dependent upon the satellite trajectories being defined on the basis of the fat ellipse. A skinny ellipse will be perceived to rotate rigidly even if the radial satellites maintain a constant distance from the contour (Weiss & Adelson, 2000). We chose to define the trajectories on the basis of the fat ellipse so that the local motions of the satellites would be the same across aspect-ratio. Thus, all conditions were perceived as rigid rotation except for the case of radial motion satellites around a fat ellipse, which was perceived to deform.

In experiment 2, the ellipses were removed from the stimulus blocks, and only the corresponding satellites were presented.

2.8. Experimental design

In each session, functional images were collected using T2* fast field echo, echo planar functional images (EPIs) sensitive to BOLD contrast: (32 axially oriented slices, TR = 1500 ms). In every run, each of the four stimulus conditions was presented in a pseudorandom order in 20-s blocks (10 TRs) separated by blank periods of 20 s during which only the fixation spot was present on the display. Each run began and finished with a blank period. A minimum of seven runs was collected per participant. As far as the subject was concerned, the sole task was to maintain fixation and press a button when the central fixation spot briefly changed color. This assured that subjects were awake, fixating, and maintaining a relatively fixed level of attention across the different experimental conditions. Runs that had fewer than 75% correct responses on this task were eliminated from further analysis.

3. Data analysis

3.1. Timecourse analysis event-related averaging

Beta weight values were computed for each stimulus condition using the general linear model (GLM) with predictors formed by convolving boxcar waveforms corresponding to the stimulus blocks with a canonical hemodynamic response function. A repeated measures ANOVA thresholded liberally at $p < 0.005$ (uncorrected) identified all visually responsive voxels. The functionally defined regions of interest described in the sections above were intersected with these voxels to create subsets corresponding to the visually responsive voxels found within each ROI. This selection criterion was used to exclude voxels corresponding to regions of retinotopically organized cortex that were mapped out using our retinotopic mapping procedures but not stimulated by the stimuli used in the main experiment. Because the experimental design included equal numbers of blocks per condition and the block-orders were randomized across run, this selection criterion was unbiased with respect to the subsequent analyses, which were limited to these visually responsive sub-ROIs.

For each participant, and each condition, event-related averages were computed for each voxel by aligning and averaging the data of corresponding conditions across runs. Each time point of the averages was computed as the percent signal deviation from baseline for the 15 TRs following the beginning of each condition. Baseline was defined as the mean response of the six TRs immediately preceding the start of a block.

3.2. Individual region of interest statistical analysis (timecourses)

For each participant and experimental condition, the event-related average was averaged across all voxels within a given ROI. The corresponding mean timecourses for a given ROI were averaged across the left and right hemispheres for each subject. From these waveforms a mean timecourse amplitude for each condition was computed by taking the mean amplitude of response from 6 to 24 s (TRs 3–12) following the start of the block. Thus for any single region of interest, the response to each of the four experimental conditions is represented by a single mean timecourse amplitude. Supplemental Figure 3 illustrates an example of the group timecourses and derived mean timecourse amplitudes for area MT.

For each region of interest, the mean timecourse amplitude for the radial motion satellite conditions was subtracted from the corresponding aspect-ratio rotational-motion satellite conditions. A satellite influence index (SII) for these differences, one each for the high and low aspect-ratio ellipses, was computed by normalizing the difference of rotation and radial conditions by their sum according to the formula $(R_{rad} - R_{rot})/(R_{rad} + R_{rot})$ to minimize the effects of between-participant variability (R_{rad} and R_{rot} represent the mean timecourse amplitude for the radial and rotational conditions, respectively).

For each region of interest, a two-tailed paired t-test was performed, contrasting the SII of high and low aspect-ratio ellipses. A significant difference within a given ROI would suggest that neuronal populations within that area differentially process the satellites and ellipses in the high and low aspect-ratio cases. Such differences likely reflect one of two aspects of the interactions between the motions of the satellites and ellipses corresponding to the two hypotheses described in Section 1. For a given ROI, a greater magnitude SII in the low (compared to the high) aspect-ratio conditions likely reflects differential processing underlying the perceptual experience of rigid and non-rigid motion. On the other hand a greater magnitude SII in the high (compared to the low) aspect-ratio conditions likely reflects differential processing underlying the segmentation of the satellite motion from the motion of the elliptical contours.

3.3. Whole brain general linear model

In addition to the ROI analyses described above, we applied the general linear model (GLM) to perform a group-based voxel-wise analysis. To do this, the data from each subject were spatially transformed into the standard Talairach space (Talairach & Tournoux, 1988). A random effects GLM analysis was then performed within each voxel in order to derive beta weights for boxcar-predictors convolved with a canonical hemodynamic response corresponding to the block presentations of each of the four experimental conditions. Thus, for every voxel, we derived beta weights for each of the four experimental conditions for each of our 11 subjects. Statistical analyses were then used to test hypotheses within each voxel. Specifically, paired t-tests were performed to identify voxels in which the radial and rotational conditions produced differential BOLD responses in both the thin and fat ellipse configurations. A repeated measures ANOVA was performed to contrast the BOLD response to the non-rigid (radial fat) percept to the BOLD responses of the three rigid percepts. Finally, a repeated measures ANOVA was performed to contrast the BOLD response to the segmented condition (radial-thin) to the BOLD responses of the three non-segmented conditions.

4. Results

4.1. Fixation

On average participants detected 90% (stderr = 2.3%) of fixation changes. Runs in which fewer than 75% of fixation changes were detected were eliminated from subsequent analyses. Due to a combination of scanner artifact and poor fixation performance, data from one participant were rejected and subsequently re-collected in a separate scanning session. Overall an average of 10 runs (min = 6, max = 12) per participant were used in the subsequent analyses.

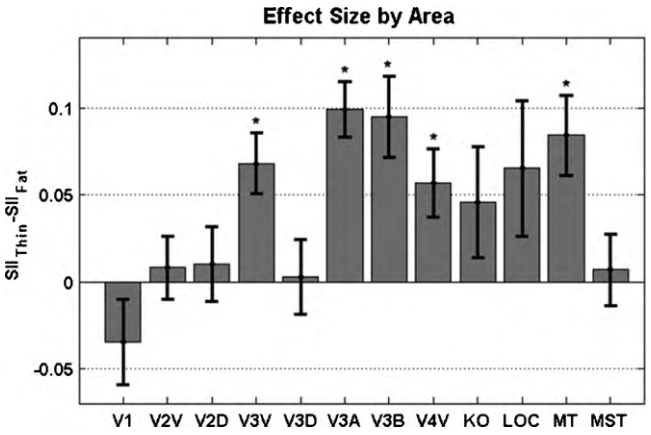

Fig. 3. Results timecourse analysis. For each area, the difference between the satellite influence index (SII) for the high and low aspect-ratio ellipses was computed using a two-tailed paired *t*-test. A positive value indicates that there was a greater difference observed between satellite conditions for the high than low aspect-ratio configurations. Significant differences were observed in V3v, V4v, V3A, V3B and MT. Error bars represent the standard error of the mean. $^{*}p < 0.05$.

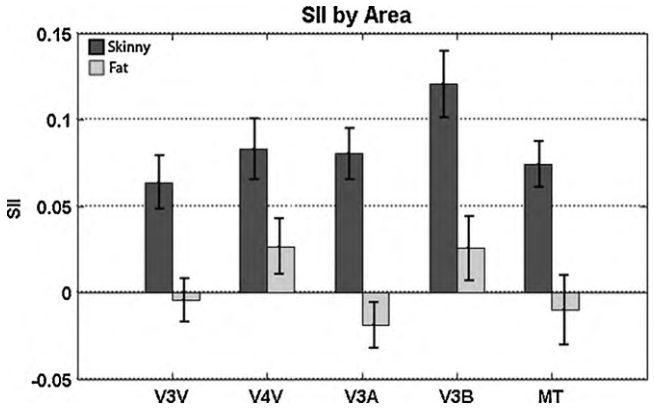

Fig. 4. Satellite influence index by area. Significant differences were observed between the SII for the high and low aspect-ratio conditions in ventral areas V3v and V4v as well as in dorsal regions V3A, V3B, and MT. In each of these areas, the SII in the high aspect-ratio condition was positive, indicating that the radial satellites led to greater BOLD signal activation than the rotational satellites. In contrast, the radial and rotational satellites led to equal BOLD signal activations in the low aspect-ratio condition. Error bars represent the standard error of the mean.

4.2. Region of interest timecourse analysis

Of particular interest are potential differences between the rotational and radial conditions for the high and low aspect-ratio ellipses. To examine these differences, a paired *t*-test was performed for each ROI contrasting the SII from the high aspect-ratio configurations with the SII from the low aspect-ratio configurations.

Fig. 3 illustrates the effect size of this contrast (mean difference between the SII_{skinny} and SII_{fat}) in each of the ROIs. Table 1 illustrates the specific statistical results of this analysis, including the Bonferroni-correct *p*-values. Across the ROIs, a significant difference in the SII between high versus low aspect-ratio ellipses was observed in V3v ($t(10) = 4.10$, $p < 0.002$), V4v ($t(10) = 3.02$, $p < 0.013$), V3A ($t(8) = 6.24$, $p < 0.001$), V3B ($t(8) = 4.07$, $p < 0.004$) and MT ($t(8) = 3.89$, $p < 0.005$). Note: the differences observed in V4v and MT were not significant at $p < 0.05$ when Bonferroni corrected for multiple comparisons. In each of these areas, the SII from the high aspect-ratio conditions was greater in magnitude than the SII from the low aspect-ratio conditions. Upon closer inspection (Fig. 4), it is interesting to note that, consistent with the predictions made for areas involved in segmentation, the radial condition in the high aspect-ratio configuration produced a larger BOLD signal response than the rotational condition in the high aspect-ratio configuration ($SII > 0$). In contrast, the BOLD signal responses to these conditions in the low aspect-ratio configuration were not significantly different ($SII \sim 0$).

Table 1
Timecourse statistical results.

	T	*p*	*p*-Bonferroni
Area			
V1	$t(10) = -1.47$	0.17	NA
V2v	$t(10) = 0.46$	0.653	NA
V2d	$t(10) = 0.48$	0.64	NA
V3v	$\mathbf{t(10) = 4.10}$	**0.002**	**0.024**
V3d	$t(10) = 0.13$	0.895	NA
V4v	$\mathbf{t(10) = 3.02}$	**0.013**	0.156
LOC	$t(8) = 1.77$	0.114	NA
KO	$t(7) = 1.35$	0.218	NA
Sub-area			
V3A	$\mathbf{t(8) = 6.24}$	**0.0002**	**0.024**
V3B	$\mathbf{t(8) = 4.07}$	**0.0036**	**0.043**
MT	$\mathbf{t(8) = 3.89}$	**0.0046**	0.055
MST	$t(8) = 0.36$	0.727	NA

The results of the paired *t*-test analysis contrasting the SII for the high and low aspect-ratio conditions. Statistically significant ($p < 0.05$) differences were observed in V3v, V4v, V3AB, V3A, V3B and MT. The positive *t*-values reflect the fact that the SII for the high aspect-ratio condition was positive and near zero in the low aspect-ratio condition. Bold formatting used to highlight statistically significant results $p < 0.05$.

A different pattern of results was observed in the control experiment where satellites were presented alone. Here, when the responses to the radial satellites were compared to the responses to the rotational satellites, no differences between conditions were observed in V1 ($t(8) = -0.734$, $p > 0.48$), V2v ($t(8) = -1.216$, $p > 0.25$), V2d ($t(8) = -1.528$, $p > 0.16$), V3v ($t(8) = -0.346$, $p > 0.73$), V3d ($t(8) = -0.380$, $p > 0.71$), V4v ($t(8) = 0.818$, $p > 0.43$), LOC ($t(7) = -1.003$, $p > 0.34$) or MST ($t(8) = -1.304$, $p > 0.22$). The comparison did reveal significant differences in areas V3A ($t(7) = -3.942$, $p < 0.006$), V3B ($t(7) = -4.339$, $p < 0.003$), KO ($t(7) = -2.489$, $p < 0.042$) and MT ($t(8) = -3.183$, $p < 0.013$). Of these ROIs, only the result in V3B yields a $p < 0.05$ when Bonferroni corrected. Based on the more liberal uncorrected *p*-values, the analyses indicate that the dorsally located V3A, V3B and MT as well as KO, respond differentially to the different satellite trajectories. This is perhaps not surprising given the fact that these areas are highly responsive to motion stimuli and the different trajectories span different areas of the visual field. The fact that the rotational trajectories lead to increased BOLD signal activity may result from the fact that the rotational trajectories traverse a greater portion of the visual field than do the radial trajectories.

However, unlike the case when the thin ellipses and satellites were both present, in this experiment greater BOLD signal activity was observed for the rotational condition as compared to the radial condition. Therefore, while V3A, V3B and MT are sensitive to differences in local motion trajectory, the differential BOLD signal activity within V3v, V4v, V3A, V3B and MT observed in the main experiment cannot be accounted for by this local processing.

4.3. Whole brain general linear model

Four distinct comparisons, as described above, were performed on the beta weights derived from the voxel-wise GLM analysis. Because, of the large number of voxels, careful consideration must be given to correcting for the large number of statistical comparisons performed. In each instance we first set a threshold corrected for a False Discovery Rate (FDR correction) of $p < 0.05$. At this conservative threshold, no significant voxels were identified under any

of the four comparisons described above. Secondarily, we selected an arbitrary threshold of an uncorrected $p < 0.0005$. The subsequent results were derived using this more liberal threshold and should be considered and interpreted in light of this fact.

The voxel-wise paired *t*-test contrasting the radial and rotational conditions in the thin ellipse configuration identified a single cluster of 138 voxels in which the responses to these two conditions differed. Shown in supplemental Figure 4, the BOLD response in these voxels, located near the inferior portion of the intraparietal sulcus in the right hemisphere (mean(std) TAL coordinates $(x,y,z) = 30(1.1), -76(1.9), 22(2.2)$) was greater in the radial-segmented condition than in the rotational conditions. While mapping spatially normalized activations onto functionally defined ROIs is imprecise, the coordinates of this cluster may correspond to what we functionally defined as V3A/B (Tootell et al., 1997) and/or V7 (Tootell et al., 1998). Indeed, 114 of 138 voxels in this cluster (83%) were found to lie in a liberally defined V3A/B mask consisting of the union of Talairached right hemisphere V3A/B masks across 57 subjects. This included individual masks from subjects run in this study and others subjects who had been retinotopically mapped for separate projects.

The paired *t*-test contrasting the responses to the radial and rotational conditions in the fat ellipse configuration revealed no significant voxels anywhere in the brain, even at the more liberal threshold of $p < 0.0005$ uncorrected. This is consistent with the ROI analyses that identified differences between the rotational and radial conditions in the thin but not fat ellipse configurations. The repeated measures ANOVA contrasting the response to the non-rigid and the responses to the rigid percepts revealed no significant voxels even at the more liberal threshold.

The repeated measures ANOVA contrasting the response to the segmented condition and the responses to the integrated conditions revealed a small cluster of 16 voxels in which the segmented condition produced a larger BOLD response. Shown in supplemental Figure 5, this cluster (mean(std) TAL coordinates $(x,y,z) = 27(0.5), -65(0.8), -7(1.1)$) is located in the posterior portion of the collateral sulcus in the right hemisphere. This location possibly corresponds to the anterior and dorsal portion of what we functionally defined as V4v that may have included portions of what others (Brewer et al., 2005) have called VO1 (for comparisons of TAL coordinates see: Arcaro et al., 2009). Indeed, 13 of 16 voxels in this cluster (81%) were found to lie in a liberally defined V4v mask consisting of the union of Talairached right hemisphere V4v masks across 52 subjects. This included individual masks from subjects run in this study and others subjects who had been retinotopically mapped. Note: the right hemisphere V4v could not be readily identified in 5 of the 57 subjects used to create the V3A/B mask described above.

5. General discussion

Certain extrastriate cortical areas show more activation when the motion information present in the image is segmented into a greater number of independent sources. In the main experiment, BOLD signals within V3v, V4v, V3A, V3B and MT were found to be differentially modulated by the two satellite trajectories as a function of aspect-ratio. The voxel-wise GLM analyses also identified a cluster of voxels in the right hemisphere that may correspond to what we functionally defined as V3A/B or perhaps the region V7 located just dorsally to V3A/B and another cluster of voxels that may correspond to the anterior and dorsal portion of what we have defined as V4v that showed a similar pattern of responses. Aside from these clusters, the voxel-wise analyses did not reveal any regions of cortex outside of those covered by the ROI analyses that showed this pattern of responses. The results of the control experiment suggest that these observed differences cannot be attributed to local differences in the radial and rotational trajectories and therefore may be attributed to the interactions between the elliptical contours and the satellites. In each of these areas, a greater difference was observed between the two satellite conditions for the skinny ellipses than for the fat ellipses as shown in Figs. 3 and 4. Specifically, in each of these areas a greater BOLD signal response was observed for the radial compared to rotational satellite condition only for the skinny ellipse. In contrast, no differences in any area were observed between the satellite motions for the fat ellipse. Thus, BOLD signals we observe do not differ with perceived rigidity/non-rigidity. As with any null result, it could be the case that differential processing within the tested ROIs does occur in response to rigid and non-rigid motions and that the methods employed here are insufficient to detect such differences. Alternatively, it could be that the differentiation of rigid and non-rigid motion occurs in areas outside those examined here. For example, the posterior portion of the superior temporal sulcus (pSTS) has been shown to respond to biological motion signals that are largely non-rigid in nature (Puce & Perrett, 2003). However, the voxel-wise general linear model analysis did not reveal any such activation. As such, our data are most consistent with the notion that the rigid and non-rigid percepts examined here are not mediated by distinct neural mechanisms but may arise as differential outcomes of a common mechanism; however, this conclusion remains speculative due to the null result from which it is derived.

It is worth mentioning that while the responses to the satellites alone (control experiment) differ in V3A, V3B, KO and MT, no such differences were observed in response to the fat ellipses presented in the main experiment. This raises the intriguing yet speculative hypothesis that in addition to mediating motion segmentation, these areas may also differentially represent rigid and non-rigid motions. Namely, it is possible that the decreased activity in response to the radial satellites alone is counteracted and obscured by increased activity in response to the non-rigid motion of the fat ellipse. However, this hypothesis rests upon an untested assumption of linearity between the responses in the main and control experiments and thus remains in the realm of speculation.

Increased BOLD signal activity found in response to the radial skinny ellipse condition presumably reflects processing involved in the segmentation of satellite and ellipse motion signals that only occurs in this condition. By 'segmentation' we mean the separation of motion signals into independent sources. This is in contrast to 'integration' in which one motion signal may influence and/or be influenced by another motion signal. Only in the radial skinny ellipse case are the dots perceived to move in a manner inconsistent with being attached to the contour by invisible but rigid rods. Thus, only in this condition is there more than one object source (the dots and the ellipse) used to account for image motion signals. The most parsimonious account of the other three conditions is that they arise from a single moving object, where the dots are in effect attached to the ellipse, even if the means by which they are attached is not visible.

These results are largely consistent with previous neuroimaging investigations of motion segmentation that found increased activity in hMT+ and other cortical areas when motion signals were segmented into transparent layers (Castelo-Branco et al., 2002; Muckli et al., 2002). The current results extend these findings to the domain of motion signals arising from rotating objects and the segmentation and integration processes associated with the disambiguation of rigid and non-rigid object motion. Furthermore, the current results specifically identify V3v and V4v as additional regions of visual cortex that are not necessarily recruited in resolving transparency in translating motion fields, yet contribute to motion segmentation in the context investigated here.

It is important to note that V3v is differentially modulated by segmentation, whereas V3d is not. To our knowledge, there is no direct neurophysiological evidence suggesting functional specialization between the ventral and dorsal regions of V3. All of the neurophysiological evidence that we are aware of suggests that the functional specifications of these regions are solely limited to differences in the quadrant of the visual field that they represent. However, recent neuroimaging studies using fMRI have reported asymmetries in ventral and dorsal activations in V1 and V2 (Caplovitz, Barroso, Hsieh and Tse, 2008; Mendola, Conner, Sharma, Bahekar, & Lemieux, 2006) and V3 (Altmann, Bulthoff, & Kourtzi, 2003; Caplovitz et al., 2008). If such a functional asymmetry exists across these subregions, one might predict that they would lead to functional asymmetries in behavior between the upper and lower visual hemifields. There has been recent debate in the literature with respect to whether such hemifield-dependent behavioral asymmetries exist (Danckert & Goodale, 2001; Hagenbeek & Van Strien, 2002; He, Cavanagh, & Intriligator, 1996; Khan & Lawrence, 2005; Previc, Breitmeyer, & Weinstein, 1995; Rubin, Nakayama, & Shapley, 1996; von Grünau & Dube, 1994; Whishaw, 1994; but see: Binsted & Heath, 2005; Krigolson & Heath, 2006). However, care must be taken when attributing functional or behavioral relevance to asymmetries observed in BOLD signals arising within a given area. Asymmetries such as those found in the data reported here between V3v and V3d could manifest themselves for reasons unrelated to the segmentation of motion signals. For example, increased BOLD signals could arise due to an increase in the spatial extent over which neural activity occurs or through increases in the magnitude of neural activity that occurs within a fixed spatial extent. Asymmetries in either the spatial distribution or relative magnitude of activity between the ventral and dorsal regions of V3 could account for the ventral/dorsal asymmetry reported here.

In order to explain the influence of satellites on the perceived motion of an ellipse, the model proposed by Weiss and Adelson (2000) suggests that if the motion-integration processes identify spatially distinct yet correlated sources of motion, then these distinct sources are combined together for purposes of global integration, as if from a common (even if disconnected) object. The local motion of both the radial and rotational satellites, for example, may be correlated to a high enough degree with the partially integrated local contour motions of a fat ellipse that they are included in the integration of motion signals arising along the entire elliptical contour. In this manner the satellites could influence the perceived rigidity of the ellipse by biasing the computation of contour motion to include radial in addition to rotational components. In the case of a skinny ellipse, partially integrated motion signals along the elliptical contour are potentially highly uncorrelated with radial satellite motions (consistent with non-rigid motion) leading to segmentation into separate 'objects' or 'motion layers', preventing the satellites from inducing a non-rigid percept. Presumably, satellite motions consistent with rotational motion are highly correlated with the integrated motion of the high aspect-ratio contour and thus get bound together by whatever grouping mechanisms underlie global integration. It is an open question just how uncorrelated the local motions of the satellites must be with the motion of the elliptical contour, before the satellites and ellipse are segmented into separate motion sources.

The factors that determine the identification, segmentation and potential integration of distinct sources of motion remain largely unknown. The above account (Weiss & Adelson, 2000) clearly posits a primary role for mechanisms within the motion processing stream and is based on the implicit assumption that independently moving objects will tend to give rise to uncorrelated motion signals. It is interesting to note that this is consistent with the case of translating dot fields. As the angle of dot motion increases between the two fields (e.g. more uncorrelated) the more likely the fields are to be segmented into transparent layers (Muckli et al., 2002). Indeed, much of the electrophysiological research investigating motion integration and segmentation (albeit largely limited to rigid translational motion) has identified area MT as playing a key role in these processes (Duncan, Albright, & Stoner, 2000; Movshon, Adelson, Gizzi, & Newsome, 1985; Pack & Born, 2001; Pack, Livingstone, Duffy, & Born, 2003; Pack, Gartland, & Born, 2004). However, the assumption that correlated motions belong to a common object need not be true of actual objects moving in the physical world. For instance, the motion of two independently moving objects can certainly be correlated, as when two dogs run together. Conversely, motions within an object can be poorly correlated, such as the spokes of a bicycle that tend to move in directions different from the direction of the bicycle itself. Furthermore, motion signals can be uncorrelated even in the absence of distinct moving objects. For example, two transparent layers of smoke or gnats can give rise to uncorrelated motion signals, even though these things are not objects. Given limitations such as these, we hypothesize that segmentation may rely upon additional information above and beyond the basic correlation of motion signals.

In the absence of object motion, the identification of distinct objects in the visual scene relies heavily upon mechanisms dedicated to the detection and processing of form features that lie outside of the motion processing stream (Mishkin, Ungerleider and Macko 1983). For example, contour features such as corners, junctions, terminators and regions of high curvature are highly informative in identifying and segmenting objects from the background (Attneave, 1954). An extensive psychophysical literature suggests that the perception of motion as a whole cannot be mediated solely on the basis of locally detected motion signals, but that local (e.g. McDermott & Adelson, 2004) and global analyses (e.g. Caplovitz & Tse, 2006, 2007b; Hsieh & Tse, 2007; Lorenceau & Alais, 2001; Tse, 2006; Tse & Caplovitz, 2006; Tse & Logothetis, 2002; Wallach, 1935) of form also contribute to the computation of object motion. It seems reasonable therefore to hypothesize that such form-processing mechanisms may also contribute to the segmentation of motion signals arising from different objects in the visual scene and the potential maintenance of object rigidity over time. Furthermore, it has been hypothesized that rather than being strictly limited to the analysis of an object's shape, aspects of an object's form can be used to influence how motion signals get integrated together (Ullman, 1979).

The recently proposed 3D "FORMOTION" model (Berzhanskaya et al., 2007) of motion perception incorporates stages of both form and motion processing and in some instances, accurately predicts the influences of satellites on the perceived rigidity of a rotating ellipse. Their model describes the response properties of neurons located within specific regions of visual cortex. Based upon these properties, the model attributes the processing of form and their interaction with the processing of motion signals to areas V1, V2, MT and MST of visual cortex. However, little direct empirical evidence exists to support the localization of the segmentation processes underlying these rigid and non-rigid percepts, and the results presented here suggest the involvement of different visual areas that should be considered in future models.

Specifically, the results of the current study suggest that in addition to the dorsally located areas V3A, V3B and MT, ventrally located areas V3v and V4v are also involved in the segmentation of the image into independent motion sources and that the perception of moving objects is supported by interactions between these areas. In support of this conclusion, a recent study found neuronal populations located in V4 that differentially respond to motion-defined contours (Mysore, Vogels, Raiguel and Orban, 2005). This suggests that the processing of relative motion signals within V4 can contribute to the analysis of an object's form. Here we extend this hypothesis and propose that form processing in V3v and V4v con-

tributes to the analysis of an object's motion. Cortical areas V3v and V4v have also been found to play a role in the form–motion interactions that underlie transformational apparent motion (Tse, 2006). The perception of transformational apparent motion specifically relies upon the segmentation and identification of distinct elements present in the image, processes that must precede the final computation of motion (Tse & Caplovitz, 2006; Tse & Logothetis, 2002). Together, these results suggest that neuronal populations within V3v and V4v may be specifically involved in the segmentation of motion information into separate objects or motion sources.

6. Conclusion

Under common viewing conditions the visual system is confronted with locally detected motion signals arising at many different locations in the visual scene. In order to construct stable and accurate percepts of the motion of objects, the visual system must segment motion signals arising from different objects and integrate motion signals arising from the same object. Understanding how the visual system resolves this problem is important for understanding motion perception in general. Here we present fMRI data suggesting that this problem is resolved within and potentially between intermediate stages of visual ventral and dorsal stream processing. The current results contribute to a growing set of neuroimaging results indicating that interactions between form and motion occur at and/or between several stages within visual cortex (Caplovitz & Tse, 2007a; Kourtzi & Kanwisher, 2000b; Krekelberg, Vatekis and Kourtzi, 2005; Tse, 2006). Specifically, activity within ventrally located V3v and V4v as well as the dorsally located areas V3A, V3B and MT reflects the form–motion interactions that mediate the segmentation and integration of motion signals arising at different locations in the visual scene. In particular, these areas respond with greater activity when the motion information present in the image has been segmented into a greater number of independent sources, even when the number of moving elements in the image is held constant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuropsychologia.2010.05.017.

References

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *30*, 523–525.

Altmann, C. F., Bulthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, *13*(4), 342–349.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *The Journal of Neuroscience*, *29*(34), 10638–10652.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61.

Berzhanskaya, J., Grossberg, S., & Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, *20*(4), 337–395.

Binsted, G., & Heath, M. (2005). No evidence of a lower visual field specialization for visuomotor control. *Experimental Brain Research*, *162*(1), 89–94.

Bonnet, C. (1981). Processing configurations of visual motion. In J. Long, & A. Bradley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, *8*(8), 1102–1109.

Burr, D. C., Baldassi, S., Morrone, M. C., & Verghese, P. (2009). Pooling and segmenting motion signals. *Vision Research*, *49*(10), 1065–1072.

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Caplovitz, G. P., & Tse, P. U. (2006). The bar-cross-ellipse illusion: Alternating percepts of rigid and non-rigid motion based on contour ownership and trackable feature assignment. *Perception*, *35*, 993–997.

Caplovitz, G. P., & Tse, P. U. (2007). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cerebral Cortex*, *17*(5), 1179–1189.

Caplovitz, G. P., & Tse, P. U. (2007 Jun). Rotating dotted ellipses: Motion perception driven by grouped figural rather than local dot motion signals. *Vision Research*, *47*(15), 1979–1991.

Caplovitz, G. P., Barroso, D. J., Hsieh, P. J., & Tse, P. U. (2008). fMRI reveals that non-local processing in ventral retinotopic cortex underlies perceptual grouping by temporal synchrony. *Human Brain Mapping*, *29*(6), 651–661.

Castelo-Branco, M., Formisano, E., Backes, W., Zanella, F., Neuenschwander, S., Singer, W., et al. (2002). Activity patterns in human motion-sensitive areas depend on the interpretation of global motion. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(21), 13914–13919.

Danckert, J., & Goodale, M. A. (2001). Superior performance for visually guided pointing in the lower visual field. *Experimental Brain Research*, *137*(3–4), 303–308.

Duncan, R. O., Albright, T. D., & Stoner, G. R. (2000). Occlusion and the interpretation of visual motion: Perceptual and neuronal effects of context. *The Journal of Neuroscience*, *20*, 5885–5897.

Fennema, C., & Thompson, W. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, *9*, 301–305.

Grzywacz, N., & Yuille, A. (1991). Theories for the visual perception of local velocity and coherent motion. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 231–252). Cambridge, MA: MIT Press.

Hagenbeek, R. E., & Van Strien, J. W. (2002). Left–right and upper–lower visual field asymmetries for face matching, letter naming, and lexical decision. *Brain and Cognition*, *49*(1), 34–44.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337.

Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge, MA: MIT Press.

Horn, B. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.

Hsieh, P. -J., & Tse, P. U. (2007). Grouping inhibits motion fading by giving rise to virtual trackable features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 57–63.

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience*, *22*(16), 7195–7205.

Khan, M. A., & Lawrence, G. P. (2005). Differences in visuomotor control between the upper and lower visual fields. *Experimental Brain Research*, *164*(3), 395–398.

Kohler, P. J., Caplovitz, G. P., & Tse, P. U. (2009). The whole moves less than the spin of its parts. *Attention, Perception & Psychophysics*, *71*(4), 675–679.

Kourtzi, Z., & Kanwisher, N. (2000a). Cortical regions involved in perceiving object shape. *The Journal of Neuroscience*, *20*(9), 3310–3318.

Kourtzi, Z., & Kanwisher, N. (2000b). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*, 48–55.

Krekelberg, B., Vatakis, A., & Kourtzi, Z. (2005). Implied motion from form in the human visual cortex. *Journal of Neurophysiology*, *94*, 4373–4386.

Krigolson, O., & Heath, M. (2006). A lower visual field advantage for endpoint stability but no advantage for online movement precision. *Experimental Brain Research*, *170*(1), 127–135.

Lorenceau, J., & Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, *4*(7), 745–751.

Lorenceau, J., & Shiffrar, M. (1992). The role of terminators in motion integration across contours. *Vision Research*, *32*, 263–273.

Marr, D. (1982). *Vision*. New York: W.H. Freeman and Co.

McDermott, J., & Adelson, E. H. (2004). Motion perception and mid-level vision. In Gazzaniga (Ed.), *The cognitive neurosciences*. MIT Press.

Mendola, J. D., Conner, I. P., Sharma, S., Bahekar, A., & Lemieux, S. (2006). fMRI Measures of perceptual filling-in in the human visual cortex. *Journal of Cognitive Neuroscience*, *18*(3), 363–375.

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: Two central pathways. *Trends in Neuroscience*, *6*, 414–417.

Movshon, J. A., Adelson, E. H., Gizzi, M., & Newsome, W. T. (1985). The analysis of moving visual patterns. In T. Chagas, R. Gattass, & C. G. Gross (Eds.), *Study group on pattern recognition* (3rd ed., pp. 117–151). Vatican City: Pontifica Academia Scientiarum.

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*(4), 843–856.

Mysore, S. G., Vogels, R., Raiguel, S. E., & Orban, G. A. (2006). Processing of kinetic boundaries in macaque V4. *Journal of Neurophysiology*, *95*(3), 1864–1880.

Nakayama, K., & Silverman, G. H. (1988a). The aperture problem – I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, *28*(6), 739–746.

Nakayama, K., & Silverman, G. H. (1988b). The aperture problem – II. Spatial integration of velocity information along contours. *Vision Research*, *28*(6), 747–753.

Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, *409*, 1040–1042.

Pack, C. C., Gartland, A. J., & Born, R. T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *The Journal of Neuroscience*, *24*, 3268–3280.

Pack, C. C., Livingstone, M. S., Duffy, K. R., & Born, R. T. (2003). End-stopping and the aperture problem: Two-dimensional motion signals in macaque V1. *Neuron*, *39*, 671–680.

Previc, F. H., Breitmeyer, B. G., & Weinstein, L. F. (1995). Discriminability of random-dot stereograms in three-dimensional space. *The International Journal of Neuroscience*, *80*, 247–253.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 435–445 (review)

Rubin, N., Nakayama, K., & Shapley, R. (1996). Enhanced perception of illusory contours in the lower versus upper visual hemifields. *Science*, *271*, 651–653.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889–893.

Slotnick, S. D., & Yantis, S. (2003). Efficient acquisition of human retinotopic maps. *Human Brain Mapping*, *18*(1), 22–29.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. Stuttgart, Germany: Thieme.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., et al. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience*, *17*(18), 7060–7078.

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., et al. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*, 1409–1422.

Tse, P. U. (2006). Neural correlates of transformational apparent motion. *Neuroimage*, *31*(2), 766–773.

Tse, P. U., & Caplovitz, G. P. (2006). Contour discontinuities subserve two types of form analysis that underlie motion processing. *Progress in Brain Research*, *154*, 271–292.

Tse, P. U., & Logothetis, N. K. (2002). The duration of 3-D form analysis in transformational apparent motion. *Perception & Psychophysics*, *64*(2), 244–265.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Vallortigara, G., Bressan, P., & Bertamini, M. (1988). Perceptual alternations in stereokinesis. *Perception*, *17*, 4–31.

Van Oostende, S., Sunaert, S., Van Hecke, P., Marchal, G., & Orban, G. A. (1997). The kinetic occipital (KO) region in man: An fMRI study. *Cerebral Cortex*, *7*(7), 690–701.

Verghese, P., & Stone, L. S. (1996). Perceived speed constrained by image segmentation. *Nature*, *381*, 61–163.

von Grünau, M., & Dube, S. (1994). Visual search asymmetry for viewing direction. *Perception & Psychophysics*, *56*, 211–220.

Wallach, H. (1935). Uber visuell Wahrgenommene Bewegungrichtung. *Psycholgishe Forschung*, *20*, 325–380.

Wallach, H., Weisz, A., & Adams, P. A. (1956). Circles and derived figures in rotation. *The American Journal of Psychology*, *69*, 48–59.

Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, *2*, 232–342.

Weiss, Y., Adelson, E. H. (1998) *Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision*. MIT AI Memo 1624 (CBCL Paper 158).

Weiss, Y., & Adelson, E. H. (2000). Adventures with gelatinous ellipses—Constraints on models of human motion analysis. *Perception*, *29*(5), 543–566.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Whishaw, I. Q. (1994). How do primates reach? *The Behavioral and Brain Sciences*, *17*, 173–174.