Behavioral/Cognitive

Integration of Motion and Form Cues for the Perception of Self-Motion in the Human Brain

Shu-Guang Kuai,[1,2] **Zhou-Kui-Dong Shan,**[2,3] **Jing Chen,**[2,3] **Zhe-Xin Xu,**[1] **Jia-Mei Li,**[1] **David T. Field,**[4] **and Li Li**[2,3]

[1]Faculty of Education, Shanghai Key Laboratory of Brain Functional Genomics, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China, [2]New York University-East China Normal University Joint Research Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai 200062, China, [3]Faculty of Arts and Science, New York University Shanghai, Shanghai 200122 China, and [4]Center for Integrative Neuroscience & Neurodynamics, Department of Psychology, University of Reading, Reading RG6 6AL, United Kingdom

When moving around in the world, the human visual system uses both motion and form information to estimate the direction of self-motion (i.e., heading). However, little is known about cortical areas in charge of this task. This brain-imaging study addressed this question by using visual stimuli consisting of randomly distributed dot pairs oriented toward a locus on a screen (the form-defined focus of expansion [FoE]) but moved away from a different locus (the motion-defined FoE) to simulate observer translation. We first fixed the motion-defined FoE location and shifted the form-defined FoE location. We then made the locations of the motion- and the form-defined FoEs either congruent (at the same location in the display) or incongruent (on the opposite sides of the display). The motion- or the form-defined FoE shift was the same in the two types of stimuli, but the perceived heading direction shifted for the congruent, but not for the incongruent stimuli. Participants (both sexes) made a task-irrelevant (contrast discrimination) judgment during scanning. Searchlight and ROI-based multivoxel pattern analysis revealed that early visual areas V1, V2, and V3 responded to either the motion- or the form-defined FoE shift. After V3, only the dorsal areas V3a and V3B/KO responded to such shifts. Furthermore, area V3B/KO shows a significantly higher decoding accuracy for the congruent than the incongruent stimuli. Our results provide direct evidence showing that area V3B/KO does not simply respond to motion and form cues but integrates these two cues for the perception of heading.

Key words: cue integration; form; heading; motion; optic flow; V3B/KO

Significance Statement

Human survival relies on accurate perception of self-motion. The visual system uses both motion (optic flow) and form cues for the perception of the direction of self-motion (heading). Although human brain areas for processing optic flow and form structure are well identified, the areas responsible for integrating these two cues for the perception of self-motion remain unknown. We conducted fMRI experiments and used multivoxel pattern analysis technique to find human brain areas that can decode the shift in heading specified by each cue alone and the two cues combined. We found that motion and form cues are first processed in the early visual areas and then are likely integrated in the higher dorsal area V3B/KO for the final estimation of heading.

Introduction

Human survival requires accurate perception and control of self-motion. How do we perceive the direction of our self-motion (heading)? Gibson (1950) proposed that humans use optic flow, a specific type of visual motion of objects in the world available at the eye generated during self-motion. When traveling on a straight path (translation), optic flow forms a radially expanding pattern and the focus of expansion (FoE) indicates our heading, in which case we can estimate heading within 1°–2° of visual angle (e.g., Warren et al., 1988; van den Berg, 1992; Crowell and Banks, 1993; L. Li et al., 2002).

Although the FoE is defined by the expanding global motion in optic flow, it is also given by global form information, such as motion streaks in a time-integrated flow field. Since Gibson's proposal, research has focused almost exclusively on what motion cues people use to perceive heading but ignored the potential influence of form cues. This could be partly due to the proposal that motion and form cues are processed by two separate visual streams that originate from the primary visual cortex and project

Received Dec. 22, 2018; revised Nov. 27, 2019; accepted Dec. 2, 2019.

Author contributions: S.-G.K. and L.L. designed research; Z.-K.-D.S., Z.-X.X., J.-M.L., and J.C. performed research; S.-G.K., Z.-K.-D.S., J.C., Z.-X.X., J.-M.L., D.T.F., and L.L. analyzed data; L.L., S.-G.K., and J.C. wrote the paper.

This work was supported by research grants from New York Universality-East China Normal University Joint Research Institute of Brain and Cognitive Science at New York University Shanghai, National Natural Science Foundation of China (31741061 and 31771209), Shanghai Science and Technology Committee (17ZR1420100 and 19JC1410101), Peak Discipline Construction Project of Education at East China Normal University, China Ministry of Education (East China Normal University 111 Project, Base B1601). We thank Qi Liang for help with data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Li Li at LL114@nyu.edu or Shu-Guang Kuai at sgkuai@psy.ecnu.edu.cn.

https://doi.org/10.1523/JNEUROSCI.3225-18.2019

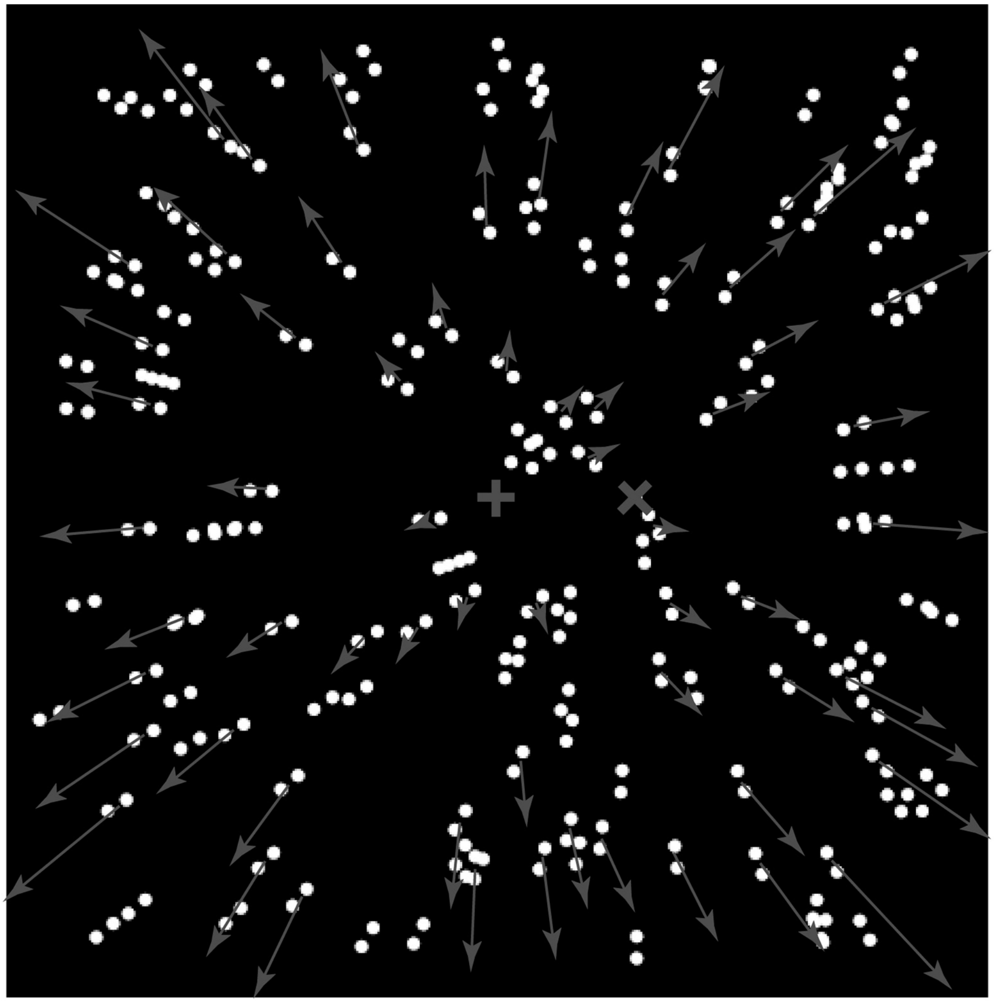
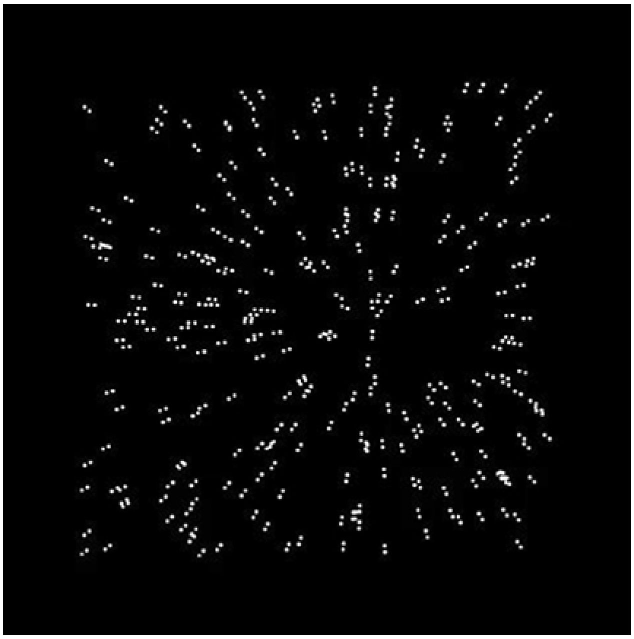


Figure 1. An illustration of the animated Glass pattern stimulus that offers two independent FoEs: the form-defined FoE given by the orientation of the dot pairs ("×") and the motion-defined FoE given by the motion of the dot pairs ("+"). Lines with arrowheads indicate velocity vectors of the centroid of the dot pairs. "×," "+," and lines with arrowheads are for illustration purpose only and not shown in the experimental stimulus.

either dorsally to the parietal cortex for motion processing or ventrally to the inferotemporal cortex for form processing (e.g., Mishkin et al., 1983; DeYoe and Van Essen, 1988).

Separate processing of motion and form information is initially supported by neuropsychological evidence from brain-damaged patients (e.g., Benson and Greenberg, 1969; Zihl et al., 1983; Goodale and Milner, 1992). However, many studies show that motion and form processing are closely linked (for review, see Kourtzi et al., 2008). For example, the classical kinetic depth effect (Wallach and O'Connell, 1953) and biological motion (Johansson, 1973) show that motion can help perceive form that could not be seen from a static display. Conversely, form can also affect motion perception, such as static "speed lines" and motion streaks depicted in diagrams are shown to bias the perceived object motion direction (e.g., Geisler, 1999; Burr and Ross, 2002).

Enlightened by these studies, Niehorster et al. (2010) developed animated Glass pattern stimuli (Glass, 1969) that pitted optic flow and form cues to self-motion against one another, with each cue indicating a different heading direction. For the first time, they found that the human visual system optimally integrates motion and form cues for heading estimation. Although the brain areas for processing motion and form cues are well identified, the areas responsible for integrating these two cues for the perception of self-motion remain unknown. To address this question, in the current study, we used similar animated Glass pattern stimuli consisting of randomly distributed dot pairs oriented toward a locus on a screen (the form-defined FoE) but moved away from a different locus (the motion-defined FoE) to simulate observer translation (Fig. 1; Movie 1). In Experiment 1, we fixed the motion-defined FoE location and shifted the form-defined FoE location. In Experiment 2, we made the locations of the motion- and the form-defined FoEs either congruent (at the same location in the display) or incongruent (on the opposite sides of the display). The shift in location of the motion- or the form-defined FoE was the same in the two types of stimuli, but the perceived direction of heading shifted in the congruent, but not in the incongruent stimuli. We performed searchlight and ROI-based multivoxel pattern analysis (MVPA) to find the brain areas that could not only respond to a location shift of the form-defined FoE (Experiment 1) but also show a higher decoding accuracy for the congruent than the incongruent stimuli (Experiment 2). These areas are likely to be in charge of integrating motion and form cues for heading perception. In Experiment 3, we randomized the form or the motion signals in the stimuli to remove the form or the motion cue to the FoE. The purpose was to validate whether the cortical areas identified in Experiments 1 and 2 are indeed driven by global form and motion signals.



Movie 1. Schematic video showing the animated Glass pattern stimuli. Dot pairs are oriented to form a radial pattern with the form-defined FoE to the right of the middle of the screen. All these dot pairs move in a radial pattern away from the motion-defined FoE in the middle of the screen.

Materials and Methods

Experimental design and statistical analyses

All experiments had within-subject design. Data were analyzed using *t* tests or repeated-measures ANOVAs with Tukey HSD tests. We reported exact *p* values. We report Cohen's *d* and η^2_p as a measure of effect size for *t* tests and ANOVAs, respectively.

Participants

Twenty-six students and staff (22 naive to the specific goals of the study) between the age of 18 and 38 years at East China Normal University and New York University Shanghai participated in the study. Among them, 14 (9 males, 5 females; mean age ± SD: 23.4 ± 5.9 years) participated in Experiment 1, 13 (9 males, 4 females; mean age ± SD: 22.8 ± 4.3 years) participated in Experiment 2, and 12 (5 males, 7 females; mean age ± SD: 23.5 ± 2.1 years) participated in Experiment 3. Fifteen participants from Experiments 1 and 2 (10 males, 5 females; mean age ± SD: 23.7 ± 5.6 years) also participated in a control psychophysical experiment.

All participants had normal or corrected-to-normal vision and provided informed consent. The study was approved by the Human Research Ethics Committee at East China Normal University and the Internal Review Board at New York University Shanghai. We determined the sample size based on the sample size in relevant previous studies.

Visual stimuli

The display simulated an observer translating at 1.5 m/s through a 3D cloud consisting of 200 white dot pairs with 0.25° centroid-to-centroid separation (dots: 0.125° in diameter, 95% luminance contrast). The 200 dot pairs were randomly placed in the depth range of 1.1–5 m according to an algorithm to ensure a uniform distribution of dot pairs as a function of distance. Dot pairs moved outside of the FOV were regenerated with the same algorithm that maintained the depth layout of the 3D cloud. In each frame, all dot pairs were oriented toward a location on the screen forming a radial Glass pattern. The display thus offered two independently generated FoEs: the form-defined FoE given by the orientation of the dot pairs (Fig. 1, "×") and the motion-defined FoE given by the centroid of dot pairs moved outward (Fig. 1, "+").

In Experiment 1, the motion-defined FoE was fixed at 0° (the center of the display) and the form-defined FoE was shifted from −5° (left) to 5° (right) by steps of 2° from the motion-defined FoE, resulting in six stimuli (see Fig. 2*a*). In Experiment 2, we tested two congruent and two incongruent stimuli. For the two congruent stimuli, the motion- and the form-defined FoEs were both at −4° or 4°. For the two incongruent stimuli, the motion- and the form-defined FoEs were at 4° on the opposite sides of the display (see Fig. 3*a*). In Experiment 3, we used the stimuli of Experiment 2 and randomized the orientation or the motion direction of the dot pairs, resulting in four form-signal-randomized stimuli and four motion-signal-randomized stimuli. Randomizing the orientation of the dot pairs removed the form-defined FoE but left the motion-defined FoE intact (see Fig. 6*a*, top row), and randomizing the motion direction of the dot pairs removed the motion-defined FoE but left the form-defined FoE intact (see Fig. 6*a*, bottom row).

On each trial, a red fixation point appeared at the center of the display for 400 ms followed by the simulated self-motion display for 600 ms. No fixation point was present in the self-motion display to ensure that the self-motion display did not contain any extraneous relative motion. Participants were instructed to fixate the fixation point that appeared at the beginning of the trial and maintain their eye position at the center of the display throughout the trial. If participants followed our instructions, then the pattern of their eye movements should not vary across the stimuli in all experiments. In 20% of trials, the contrast of half of the dot pairs was lowered by ∼50%. Participants were asked to watch the display carefully and press a button to report the trials containing dots with lower contrast.

To examine heading perception with the congruent and the incongruent stimuli, we conducted a control psychophysics experiment. For the two congruent stimuli, the motion- and the form-defined FoEs were in the same location that was randomly sampled from −3° (left) to 3° (right) by steps of 0.5° (i.e., 13 locations) with respect to a vertical reference line. The reference line was located at −4° or 4° with respect to the center of the display. For the two incongruent stimuli, the reference line was located at the center of the display. The motion- and the form-defined FoEs were 8° apart on the opposite sides of the display. The location of the motion-defined FoE was randomly sampled from −7° to −1° or 1° to 7° by steps of 0.5° (i.e., 13 locations) with respect to the reference line. Similar to the brain-imaging experiment, on each trial, a white fixation cross appeared at the center of the display for 400 ms followed by the simulated self-motion for 600 ms. Participants were instructed to fixate the fixation cross that appeared at the beginning of the trial and maintain their eye position at the center of the display throughout the trial. Right after the motion, the vertical reference line (blue, 0.8° H) appeared along the azimuth of the display, and participants were asked to press a mouse button to indicate whether their perceived direction of heading was to the left or right of the reference line. To prevent participants from memorizing the location of the reference line, its position was jittered in the range of −1° to 1° in each trial. We fitted a cumulative Gaussian function to participants' heading judgment data. The mean of the fitted Gaussian function indicates the point of subjective equality (PSE) in heading judgments (i.e., the perceived direction of heading).

Equipment and imaging acquisition parameters

The display was rendered with Psychtoolbox-3 Toolbox and back-projected on a white screen (resolution: 1024H × 768V pixels; refresh rate: 60 Hz) in a Siemens 3T MRI scanner (Trio for Experiments 1 and 2, Prisma for Experiment 3). Participants lay supine in the scanner and viewed the display (19° × 19°) binocularly through light-reflecting mirrors at the distance of 92 cm. Participants' head was positioned in a 32-channel head coil for enhanced signal to noise. Functional scans consisted of repeated EPI: voxel size = 3 × 3 × 3 mm, TE = 30 ms, flip angle = 81°, matrix size = 64 × 64, FOV = 192 × 192 mm, 38 slices (0.3 mm interslice gap, 3.0 mm slice thickness, slice order interleaved ascending), and TR = 2000 ms. A detailed T1-weighted anatomical image was also acquired: voxel size = 1 × 1 × 1 mm, TE = 2.34 ms, flip angle = 7°, FOV = 256 × 256 mm, 192 slices (no gap), and TR = 2530 ms (5 min and 48 s total scan time).

In the psychophysics experiment, the display was presented on an ASUS VG278H 27-inch LCD monitor (resolution: 1024H × 768V pixels; refresh rate: 60 Hz). Participants viewed the display (19° × 19°) binocularly with their head stabilized by a chin rest at 57 cm away from the display.

Procedure

In all three experiments, participants were scanned for eight runs using a block design. Each run had 24 stimulus blocks (6 stimuli × 4 blocks) in Experiment 1, 16 stimulus blocks (4 stimuli × 4 blocks) in Experiment 2, and 24 stimulus blocks (8 stimuli × 3 blocks) in Experiment 3. Each 16-s stimulus block contained 16 trials of a stimulus. The testing order of stimulus was randomized in each run. No feedback was provided in any stimulus trial. Each run also had a 16-s fixation block with no stimulus but a red fixation point in the center of a blank screen at the beginning, in the middle, and at the end of the run. The purpose of the fixation block was to acquire baseline brain activations in each run. The scanning lasted about 1 h for Experiment 1, about 40 min for Experiment 2, and about 1 h for Experiment 3.

For each participant in a separate scanning session that lasted about 1 h, we identified the following ROIs: the early visual areas that respond to both local motion and form information (V1, V2), the higher ventral areas that respond to shape and global form information (V3v, hV4, LO), the dorsal (hMST) and the parietal areas (VIP, V6), and area CSv that respond to optic flow. Because previous human brain-imaging studies have shown that the dorsal stream can be activated by both motion and form information (Braddick et al., 2000; Krekelberg et al., 2005), we also identified other visual areas along the dorsal stream (V3d, V3a, V7, V3B/KO, hMT) that are known to respond to motion information. Specifically, we identified the retinotopic visual areas (V1, V2, V3v, V3d, V3a, hV4, V7) using standard retinotopic mapping procedures with rotating wedge stimuli (Engel et al., 1994; Sereno et al., 1995; DeYoe et al., 1996). We defined area hV4 as the ventral, but not the dorsal, subregion of V4 (Wandell et al., 2007). We identified areas V3B/KO (DuPont et al., 1997; Zeki et al., 2003), LO (Kourtzi and Kanwisher, 2001), hMT (Zeki et al., 1991), hMST (Dukelow et al., 2001), V6 (Pitzalis et al., 2010), and CSv (Wall and Smith, 2008) using independent localizers as described in the cited studies. Finally, we identified area VIP (average center of ROI: −26, −64, 43 for left and 30, −61, 46 for right) by comparing the anatomical structure of the activated areas in the experiments to what is described in previous studies (e.g., Orban et al., 2004, 2006).

To examine whether participants could follow the instructions to fixate the fixation point that appeared at the center of the display at the beginning of a trial and then maintain their eye position there throughout the trial, in a separated session outside of the scanner, we recorded eye movements of 6 randomly selected participants from Experiments 1 and 2 (3 males, 3 females; mean age ± SD: 22.5 ± 2.3 years) using an Eyelink 1000 plus eye tracker (1 kHz, SR Research) when they viewed the same display (19° × 19°) on an LCD monitor (1024 × 768 pixels, 60 Hz) and performed the same task as in Experiments 1 and 2.

In the psychophysics experiment, each participant completed a total of 260 experimental trials (4 stimulus conditions × 13 FoE combinations × 5 trials). The trials were blocked by stimulus condition and randomized within each block. The testing order of stimulus condition was counterbalanced between participants. Participants received 5–10 practice trials at the beginning of each block. No feedback was provided in any practice or experimental trial. The psychophysics experiment lasted about 30 min.

Data analysis

Preanalysis. Neuroimaging data were analyzed using Brain Voyager QX (Brain Innovations). The anatomical data were transformed into the standard MNI space and then inflated using BrainVoyager QX. Preprocessing of the functional data included slice scan time correction, 3D motion correction, linear trend removal, and temporal high-pass filtering. The EPI images were then aligned with the anatomical images and transformed into the standard MNI space. All functional data were transformed into a 3 mm isovoxel volume time course data using the nearest neighbor algorithm without spatial smoothing.

MVPA. We performed MVPA (Haynes and Rees, 2005; Kamitani and Tong, 2005) to decode BOLD responses evoked by different stimuli. We first normalized the time course data by computing the *Z* scores of BOLD signals in each run to minimize the baseline difference across runs. We shifted the time course data forward by 4 s to compensate for the hemodynamic response delay and then averaged the BOLD response data across trials in each stimulus block. For the ROI-based analysis, we conducted a GLM analysis to select the most activated voxels in each ROI by comparing their responses in the stimulus blocks with their baseline responses in the fixation blocks. For the searchlight analysis (Kriegeskorte et al., 2006), we defined a spherical aperture (radius: 9 mm) and moved this aperture voxel by voxel across the gray matter of each participant's brain where the responses in the stimulus blocks were higher than in the fixation blocks. We then trained a linear support vector machine (SVM) classifier to discriminate the selected voxels' BOLD responses to different stimuli in the stimulus blocks of randomly selected training runs. After training, we then calculated the accuracy of the classifier's prediction of the stimuli in the stimulus blocks in the testing runs. We repeated this procedure many times for cross validation. The mean prediction accuracy of the stimuli in the testing runs averaged across the repetition times was defined as the classifier's decoding accuracy.

To estimate the significance level of the classifier's decoding accuracy, we performed a shuffled analysis in which we randomly assigned the stimulus labels to the stimuli in the stimulus blocks during training. We then conducted the same prediction procedure using the stimulus blocks in the testing runs and repeated it for 1000 times. The mean prediction accuracy of the stimuli in the testing runs averaged across 1000 times was defined as the classifier's baseline decoding accuracy.

Results

Areas encoding form-defined FoEs

Experiment 1 was designed to find the human brain areas that respond to a shift in location of the form-defined FoE (i.e., encode form-defined FoEs). Specifically, we fixed the motion-defined FoE at the center of the display (0°) and shifted the location of the form-defined FoE from −5° (left) to 5° (right) by steps of 2°, resulting in six stimuli (Fig. 2*a*). For each ROI, we thus trained a six-way linear SVM classifier to discriminate the pattern of BOLD response to the six stimuli using the data from three pairs of even- and odd-numbered runs randomly selected from the eight experimental runs. The classifier's prediction accuracy of the six stimuli was calculated using the data from the remaining pair of even- and odd-numbered experimental runs. This procedure was repeated 16 times to cover all possible combinations of even- and odd-numbered runs for training and testing. Fig. 2-1 (available at https://doi.org/10.1523/JNEUROSCI.3225-18.2019.f2-1) plots the classifier's mean prediction accuracy (i.e., decoding accuracy) of the six stimuli averaged across 16 repetitions as a function of the number of the most activated voxels (starting from 50 to the minimum number of the activated voxels across all participants) for each ROI. For all ROIs, the decoding accuracy is stabilized with the voxel number of ≥100. Figure 2*b* (gray bars) plots the classifier's decoding accuracy with 100 most activated voxels for each ROI.

We grouped the ROIs as the early visual areas (V1, V2), the ventral visual areas (V3v, hV4, LO), the dorsal visual areas (V3d, V3a, V7, V3B/KO), the dorsal motion visual areas (hMT, hMST), and other optic flow areas (VIP, V6, CSv). We conducted a two-way (ROI × decoding vs baseline decoding accuracy) repeated-measures ANOVA for each group and found that the classifier's decoding accuracy was significantly higher than its baseline decoding accuracy for the early ($F_{(1,13)} = 8.91$, $p = 0.011$, $\eta_p^2 = 0.41$), the ventral ($F_{(1,13)} = 7.86$, $p = 0.015$, $\eta_p^2 = 0.38$), and the dorsal ($F_{(1,13)} = 7.13$, $p = 0.019$, $\eta_p^2 = 0.35$) visual areas. No such main effect was found for the dorsal motion visual areas ($F_{(1,13)} = 1.22$, $p = 0.29$, $\eta_p^2 = 0.086$) or other optic flow areas ($F_{(1,13)} =$

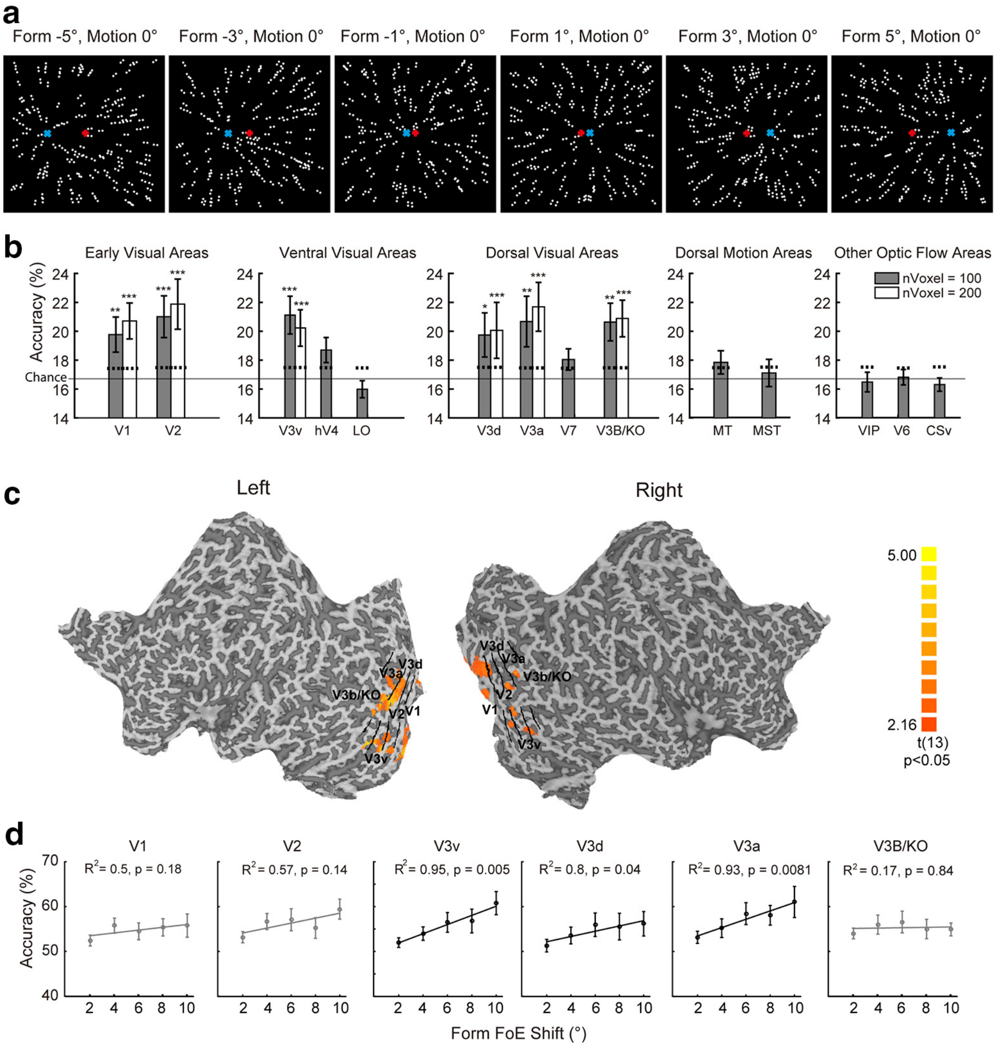


Figure 2. Experiment 1 visual stimuli and data. ***a***, Illustrations of the six stimuli. Negative sign indicates the FoE location to the left of the display center; positive sign indicates the FoE location to the right. "×" and "+" indicate the form- and the motion-defined FoEs, respectively. ***b***, The classifier's decoding accuracy for the six stimuli for each ROI group. Gray bars plot the data with 100 most activated voxels. White bars plot the data with 200 most activated voxels. (see also Figure 2–1, available at https://doi.org/10.1523/JNEUROSCI.3225-18.2019.f2-1) Dotted lines indicate the upper limits of the 95% CI of the classifier's baseline decoding accuracies from 1000 shuffled tests. Solid line indicates the chance level of 1/6. Error bars indicate SEs across 14 participants. ***$p < 0.001$. **$p < 0.01$. *$p < 0.05$. ***c***, The searchlight brain map showing clusters (≥25 voxels) that have significantly higher decoding accuracies than the baseline levels across 14 participants ($t_{(13)} > 2.16, p < 0.05$). ***d***, The classifier's decoding accuracy as a function of the form-defined FoE position shift. Solid lines indicate the fitted linear functions. Error bars indicate SEs across 14 participants.

0.18, $p = 0.68$, $\eta_p^2 = 0.014$). Tukey HSD tests revealed that the classifier's decoding accuracy was significantly higher than its baseline decoding accuracy for areas V1 ($p = 0.0016$), V2 ($p = 0.00025$), V3v ($p = 0.00029$), V3d ($p = 0.020$), V3a ($p = 0.0010$), and V3B/KO ($p = 0.0012$), indicating that the pattern of BOLD response of these visual areas can be modulated by the shift in location of the form-defined FoE in the display. Because the minimum number of the activated voxels across participants was >200 for these six areas (see Fig. 2-1, available at https://doi.org/10.1523/JNEUROSCI.3225-18.2019.f2-1), we thus also computed the classifier's decoding accuracies by selecting 200 most activated voxels in these areas as plotted in Figure 2*b* (white bars). We then conducted a 2 (100 vs 200 voxels) × 2 (decoding vs baseline decoding accuracy) repeated-measures ANOVA for each of these areas. Tukey HSD tests revealed that same as with 100 voxels, with 200 voxels, the classifier's decoding accuracy was also significantly higher than its baseline decoding accuracy for all the six visual areas (V1: $p = 0.00020$; V2: $p = 0.00020$; V3v: $p = 0.00025$; V3d: $p = 0.00027$; V3a: $p = 0.00020$; V3B/KO: $p = 0.00020$). Because the classifier's decoding reliability in general increases with the number of voxels (Haynes and Rees, 2005; Furlan et al., 2014), we trained the classifier and computed its decoding accuracy by selecting 200 most activated voxels for these six areas in the following analyses.

To examine whether any high-level brain areas also respond to the form-defined FoE shift, we conducted searchlight MVPA analysis. The classifier's decoding accuracy was computed for the central voxel of each spherical aperture, resulting in a map of decoding accuracy of the whole brain for each participant. We set a cluster size threshold of 25 voxels and performed one-sample *t* tests to compare each cluster's decoding accuracy with the chance level of 1/6 across participants. We found that, consistent with the results of the ROI-based MVPA analysis, the early visual areas V1 and V2, the ventral visual area V3v, and the dorsal visual areas V3d, V3a, and V3B/KO showed significantly higher decoding accuracies than the baseline levels. Furthermore, we did not observe any high-level brain areas involved in decoding the form-defined FoE shift (Fig. 2*c*).

How do these brain areas represent form-defined FoEs in heading perception? Do they respond only to the form-defined FOE position shift or their response can be modulated by the magnitude of the position shift? To address this question, we trained a two-way linear SVM classifier with 200 voxels to discriminate the pattern of BOLD response when the relative position shift of the form-defined FoE in two of the six stimuli was 2°, 4°, 6°, 8°, or 10°. Figure 2*d* plots the classifier's decoding accuracy as a function of the form-defined FoE position shift for these brain areas. A simple linear regression analysis revealed a significant linear trend between the decoding accuracy and the form-defined FoE position shift for areas V3v ($R^2 = 0.95$, $p = 0.0050$), V3d ($R^2 = 0.80$, $p = 0.040$), and V3a ($R^2 = 0.93$, $p = 0.0081$) but not for areas V1 ($R^2 = 0.50$, $p = 0.18$), V2 ($R^2 = 0.57$, $p = 0.14$), and V3B/KO ($R^2 = 0.17$, $p = 0.84$). This suggests that, while all these six areas respond to the form-defined FOE position shift, only the responses in areas V3v, V3d, and V3a can be modulated by the magnitude of the position shift.

In summary, this experiment allowed us to identify the brain areas that respond to a shift in location of the form-defined FoE. We found that the pattern of BOLD response in areas V1, V2, V3v, V3d, V3a, and V3B/KO changed with the shift in location of the form-defined FoE. Because the motion-defined FoE was fixed in all six stimuli in this experiment, it remains in question whether these areas also respond to a shift in location of the motion-defined FoE, and if so, how these areas integrate motion and form signals for the perception of heading. Experiment 2 was designed to address these questions.

Areas integrating motion and form cues for heading perception

In Experiment 2, we tested two types of stimuli in which the form- and the motion-defined FoE locations were congruent (i.e., both were at −4° or 4°) or incongruent (i.e., the motion-defined FoE was at −4° and the form-defined FoE was at 4° or vice versa; Fig. 3*a*). Before scanning, we conducted the psychophysics experiment to examine participants' heading perception. We found that, for the two congruent stimuli, the mean PSE averaged across 15 participants was $-4.34 \pm 0.15°$ (mean ± SE) or $4.61 \pm 0.18°$ when the motion- and the form-defined FoEs were at −4° or 4°. For the two incongruent stimuli, the mean PSE was $-0.24 \pm 0.59°$ or $-0.66 \pm 1.21°$ when the motion-defined FoE was at 4° and the form-defined FoE was at −4° or vice versa (Fig. 3*b*). Separate paired *t* tests revealed that the mean PSE was significantly different for the two congruent stimuli ($t_{(14)} = -50.84$, $p < 0.0010$, Cohen's $d = -13.13$) but not for the two incongruent stimuli ($t_{(14)} = 0.15$, $p = 0.89$, Cohen's $d = 0.038$). This indicates that the perceived direction of heading shifted with the congruent but not with the incongruent stimuli. Due to the fact that the changes in motion and form signals in the two congruent stimuli were the same as in the two incongruent stimuli, the brain areas that show a higher decoding accuracy for the congruent than the incongruent stimuli should be responding to the perceived direction of heading rather than the change in motion or form signals.

Following this logic, we trained a two-way linear SVM classifier to discriminate the pattern of BOLD response to the two congruent or incongruent stimuli using the data from three pairs of even- and odd-numbered runs randomly selected from the eight experimental runs. The classifier's prediction accuracy of the two congruent or incongruent stimuli was calculated using the data from the remaining pair of even- and odd-numbered experimental runs. This procedure was again repeated 16 times to cover all possible combinations of even- and odd-numbered runs for training and testing. Figure 3*c* plots the classifier's mean prediction accuracy (i.e., decoding accuracy) of the two congruent or incongruent stimuli averaged across 16 repetitions for the six visual areas identified in Experiment 1. A 2 (decoding vs baseline decoding accuracy) × 2 (congruent vs incongruent stimuli) repeated-measures ANOVA revealed that, for areas V1, V2, V3v, V3d, and V3a, only the main effect of decoding accuracy was significant ($F_{(1,12)} > 22.03$, $p < 0.00052$, $\eta_p^2 > 0.64$). For area V3B/KO, both the main effects of decoding accuracy and stimulus type as well as their interaction effect were significant ($F_{(1,12)} = 21.32$, $p = 0.00060$, $\eta_p^2 = 0.64$, $F_{(1,12)} = 6.63$, $p = 0.024$, $\eta_p^2 = 0.36$, and $F_{(1,12)} = 6.72$, $p = 0.024$, $\eta_p^2 = 0.36$, respectively). Tukey HSD tests showed that the classifier's decoding accuracy for the two congruent or incongruent stimuli was significantly higher than its corresponding baseline level for all the six visual areas ($p < 0.0092$), indicating that these areas can discriminate the two stimuli of either the congruent or incongruent type. Nevertheless, while there was no significant difference in the classifier's decoding accuracy between the congruent and the incongruent stimuli for areas V1 ($p = 0.997$), V2 ($p = 0.52$), V3v ($p = 0.86$), V3d ($p = 0.67$), and V3a ($p = 0.38$), the classifier's decoding accuracy was significantly higher for the congruent than the incongruent stimuli for area V3B/KO ($p = 0.015$). This suggests that area V3B/KO plays an important role in the integration of motion and form signals for the perception of heading.

Both the motion- and the form-defined FoEs changed their locations in the two congruent and the two incongruent stimuli. The higher-than-baseline decoding accuracy observed for both types of stimuli thus does not tell us whether the brain area responded to a shift in location of the motion- or the form-defined FoE or both. To separate the brain area's responses to the motion- and the form-defined FoE shifts, we examined the BOLD responses to the stimuli in which the shift in location only happened for the motion- or the form-defined FoE (Fig. 4*a*). To illustrate, in the stimuli when only the location of the motion-defined FoE was shifted, the form-defined FoE was fixed (at −4° or 4°) while the motion-defined FoE was shifted from −4° to 4°. Similarly, in the stimuli when only the location of the form-defined FoE was shifted, the motion-defined FoE was fixed (at −4° or 4°) while the form-defined FoE was shifted from −4° to 4°. We thus trained a two-way linear SVM classifier to discriminate the pattern of BOLD response to the motion- or the form-defined FoE shift using the data from three pairs of even- and odd-numbered runs randomly selected from the eight experimental runs. The classifier's prediction accuracy of the motion- or the form-defined FoE shift was calculated using the data from the remaining pair of even- and odd-numbered experimental runs. This procedure was again repeated 16 times to cover all possible combinations of even- and odd-numbered runs for training and testing.

Figure 4*b* plots the classifier's mean prediction accuracy (i.e., decoding accuracy) of the motion- or the form-defined FoE shift averaged across 16 repetitions for the six visual areas identified in Experiment 1. A 2 (decoding vs baseline decoding accuracy) × 2 (form vs motion cue) repeated-measures ANOVA revealed that, while the main effect of decoding accuracy was significant for all six visual areas ($F_{(1,12)} > 13.00$, $p < 0.0037$, $\eta_p^2 > 0.52$), the main effect of cue type and the interaction effect of decoding accuracy and cue type were also significant for areas V1 ($F_{(1,12)} = 12.73$, $p = 0.0039$, $\eta_p^2 = 0.52$ and $F_{(1,12)} = 12.90$, $p = 0.0037$, $\eta_p^2 = 0.52$, respectively) and V2 ($F_{(1,12)} = 5.90$, $p = 0.032$, $\eta_p^2 = 0.33$ and $F_{(1,12)} = 6.23$, $p = 0.028$, $\eta_p^2 = 0.34$, respectively).

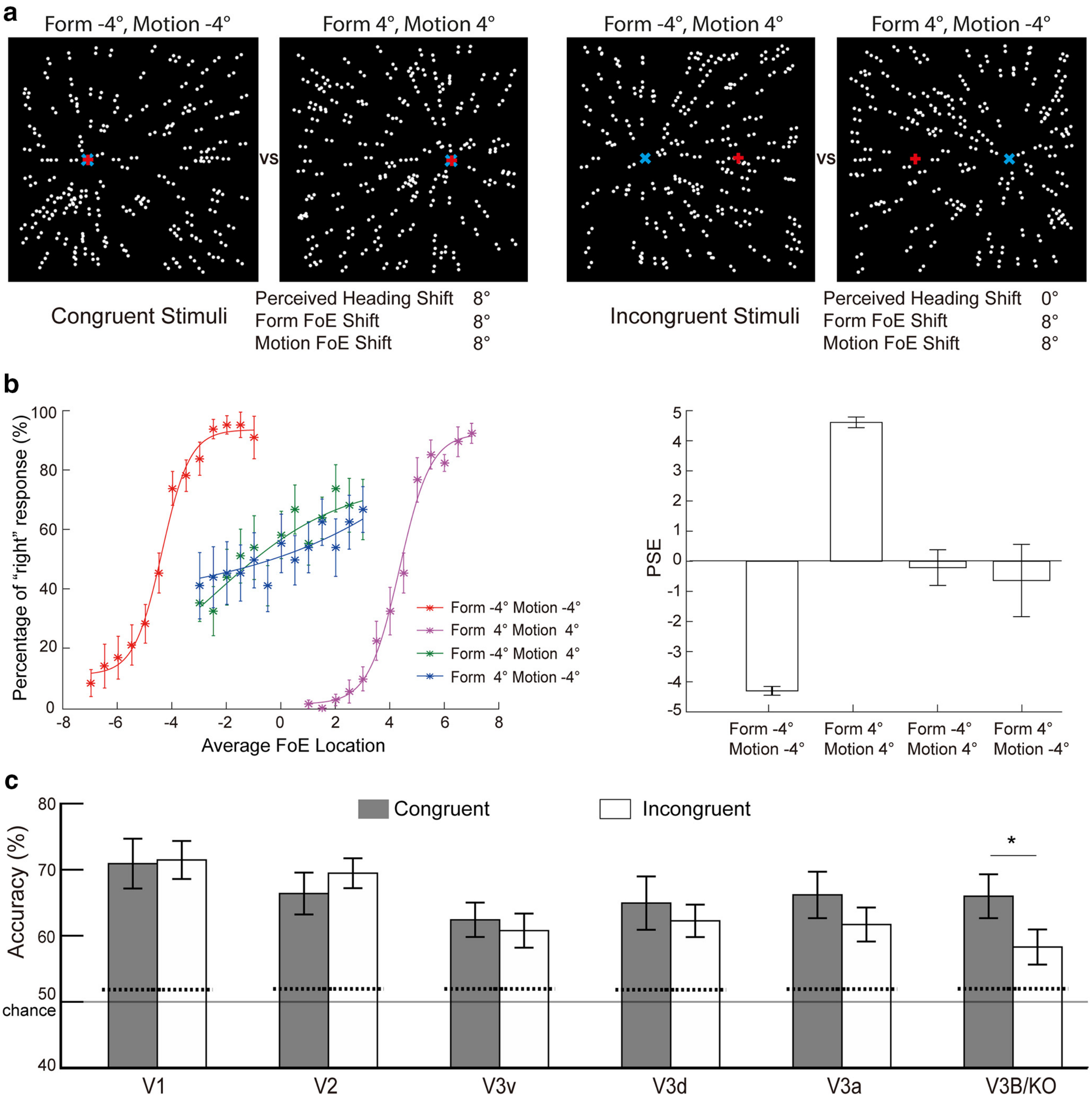


Figure 3. Experiment 2 visual stimuli and data. ***a***, Illustrations of the four stimuli. Negative sign indicates the FoE location to the left of the display center; positive sign indicates the FoE location to the right. "×" and "+" indicate the form- and the motion-defined FoEs, respectively. ***b***, Data from the psychophysics experiment. Left: Mean percentage of "right" response in heading judgments as a function of the average location of the motion- and the form-defined FoEs. Solid lines indicate cumulative Gaussian functions fitted to the data averaged across participants. Right: Mean PSE against stimulus type. Error bars indicate SEs across 15 participants. ***c***, The classifier's decoding accuracy for the congruent (gray) and incongruent (white) stimuli for the six visual areas that encode the form-defined FoE shift as identified in Experiment 1. Dotted lines indicate the upper limits of the 95% CI of the classifier's baseline decoding accuracies from 1000 shuffled tests. Solid line indicates the chance level of 1/2. Error bars indicate SEs across 13 participants. *$p < 0.05$.

Tukey HSD tests revealed that, for area V1, the decoding accuracy for the motion-defined FoE shift was significantly higher than the baseline level ($p = 0.00022$) and the decoding accuracy for the form-defined FoE shift was borderline significantly higher than the baseline level ($p = 0.085$). For all the other five areas (V2, V3v, V3d, V3a, and V3B/KO), the decoding accuracy for either the motion- or the form-defined FoE shift was significantly higher than the corresponding baseline level ($p < 0.039$), indicating that these areas respond to either the motion- or the form-defined FoE shift. For areas V1 and V2, the decoding accuracy was also significantly higher for the motion- than the form-defined FoE shift ($p < 0.020$), indicating that these two areas have a higher sensitivity to the motion than the form cue to the FoE.

Neural computation for integrating motion and form cues

How do the brain areas that encode either the motion- or the form-defined FoE shift combine motion and form cues when they are both present? There are two possibilities: linear optimal combination and fusion. For linear optimal combination, the

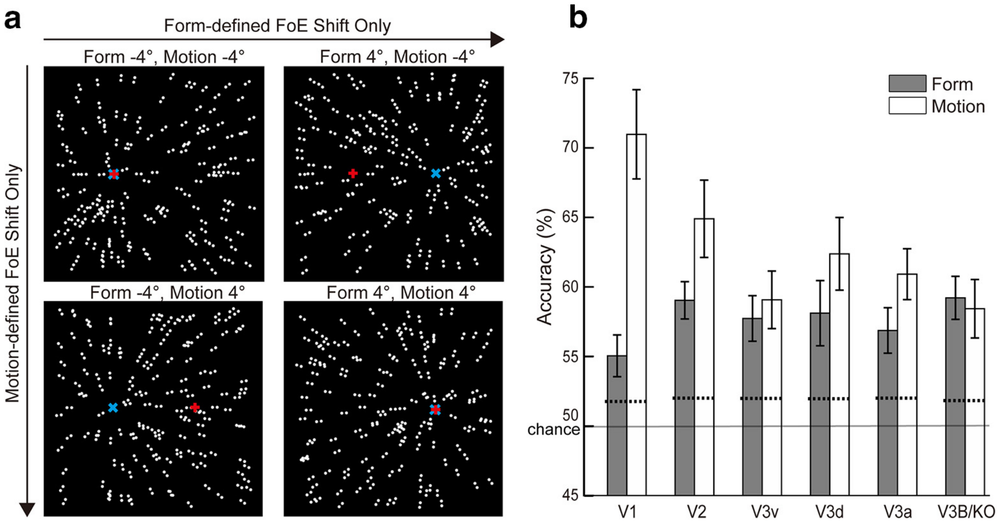


Figure 4. Visual stimuli and decoding accuracies for the motion- or the form-defined FoE shift. ***a***, Illustrations of the stimuli with only the motion- or only the form-defined FoE shift. "×" and "+" indicate the form- and the motion-defined FoEs, respectively. ***b***, The classifier's decoding accuracy for the motion- (white) or the form-defined FoE shift (gray) for the six visual areas identified in Experiment 1. Dotted lines indicate the upper limits of the 95% CI of the classifier's baseline decoding accuracies from 1000 shuffled tests. Solid line indicates the chance level of 1/2. Error bars indicate SEs across 13 participants.

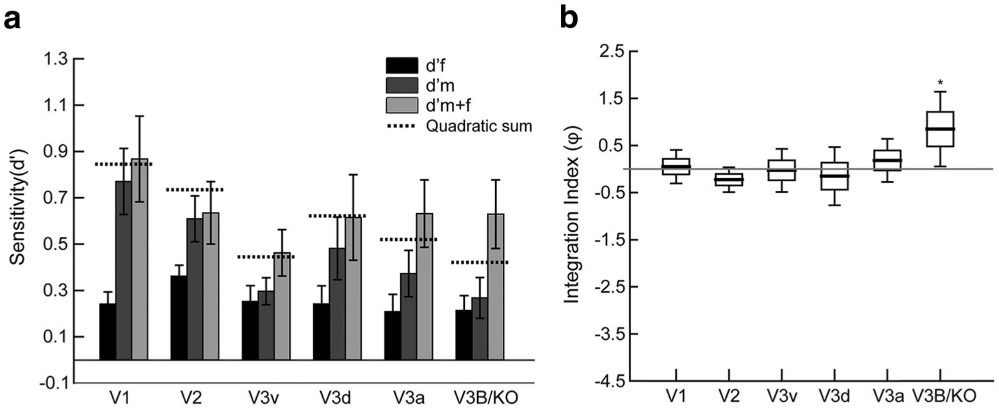


Figure 5. Sensitivity and integration index data. ***a***, The motion cue (d'_m), the form cue (d'_f), and the combined cue (d'_{m+f}) sensitivity indices for the six visual areas. Dotted lines indicate the quadratic sums of d'_m and d'_f. Error bars indicate SEs across 13 participants. ***b***, The integration indices for the six visual areas. Black line in the center of each bar indicates the median. Edges represent 68% CIs. Error bars indicate the 95% CIs. *$p < 0.05$.

brain area processes two types of cues as independent components and combines them in a statistically optimal manner according to the Bayes theorem (Landy et al., 1995; Ban et al., 2012). In this case, the classifier's decoding sensitivity to two consistent cues should be the quadratic sum of its sensitivity to each cue alone. In contrast, for fusion, the brain area may not process the two cues independently and may combine them in a nonlinear way (Ban et al., 2012). In this case, the classifier's decoding sensitivity to two consistent cues would not be equal to the quadratic sum of its sensitivity to each cue alone.

A classifier's sensitivity index (d') to decode the neural responses to a cue can be computed using its decoding accuracy for that cue (Ban et al., 2012), given as:

$$d' = 2erf^{-1}(2p - 1) \tag{1}$$

where p is the decoding accuracy. To examine how brain areas combine motion and form cues for heading perception, we computed the classifier's form cue decoding sensitivity index (d'_f) using its decoding accuracy for only the form-defined FoE shift (Fig. 4*b*, gray bars), the classifier's motion cue decoding sensitivity index (d'_m) using its decoding accuracy for only the motion-defined FoE shift (Fig. 4*b*, white bars), and the classifier's combined cue decoding sensitivity index (d'_{m+f}) using its decoding accuracy for both the motion- and the form-defined FoE shift in the two congruent stimuli (Fig. 3*c*, gray bars). Figure 5*a* plots d'_f, d'_m, d'_{m+f}, and the quadratic sum of d'_m and d'_f for the six visual areas identified in Experiment 1. To make the comparison of d'_{m+f} to the quadratic sum of d'_m and d'_f easier, we converted the sensitivities indices to an integration index (φ) as follows:

$$\varphi = \frac{d'_{m+f}}{\sqrt{d'^2_f + d'^2_m}} - 1 \tag{2}$$

When d'_{m+f} is equal to the quadratic sum of d'_m and d'_f, the integration index φ would be zero. Figure 5*b* plots the integration index for the six visual areas. Separate *t* tests revealed that, while the integration index was significantly larger than zero for area V3B/KO ($t_{(12)} = 2.31$, $p = 0.040$, Cohen's $d = 0.64$), it was not significantly different from zero for the other five areas (V1: $t_{(12)} = 0.30$, $p = 0.77$, Cohen's $d = 0.083$; V2: $t_{(12)} = -1.86$, $p = 0.088$, Cohen's $d = -0.52$; V3v: $t_{(12)} = -0.13$, $p = 0.90$; V3d: $t_{(12)} = -0.53$, $p = 0.61$, Cohen's $d = -0.15$; V3a: $t_{(12)} = 0.86$, $p = 0.41$, Cohen's $d = 0.24$). This suggests that, in contrast to areas V1, V2, V3v, V3d, and V3a that perform linear optimal combination when responding to motion and form cues, area V3B/KO performs fusion computation when combining these two cues for heading perception.

Randomizing form or motion signals

To validate whether the responses in the brain areas identified in Experiments 1 and 2 are indeed driven by global form and motion signals, in Experiment 3, we randomized the form signals in the four stimuli of Experiment 2 by randomizing the orientation of the dot pairs or the motion signals by randomizing the motion direction of the dot pairs, resulting in eight stimuli. Randomizing the form signals removed the form-defined FoE in the display but left the motion-defined FoE intact (Fig. 6*a*, top row), and randomizing the motion signals removed the motion-defined FoE but left the form-defined FoE intact (Fig. 6*a*, bottom row). Despite the randomization of the form or motion signals in the stimuli, we kept the label of each stimulus condition the same as in Experiment 2 for the data analysis and comparison purposes.

As in Experiment 2, we trained a two-way linear SVM classifier to discriminate the pattern of BOLD response to the motion- or the form-defined FoE shift using the data from three pairs of even- and odd-numbered runs randomly selected from the eight experimental runs. The classifier's prediction accuracy of the motion- or the form-defined FoE shift was calculated using the data from the remaining pair of even- and odd-numbered experimental runs. This procedure was repeated 16 times to cover all possible combinations of even- and odd-numbered runs for training and testing. Figure 6*b* plots the classifier's mean prediction accuracy (i.e., decoding accuracy) of the motion- or the form-defined FoE shift averaged across 16 repetitions for the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right) for the six visual areas. A 2 (decoding vs baseline decoding accuracy) × 2 (form vs motion cue) repeated-measures ANOVA revealed that for both the form- and the motion-signal-randomized stimuli, the interaction effect of decoding accuracy and cue type was significant for all six visual areas ($F_{(1,11)} > 6.55$, $p < 0.027$, $\eta_p^2 > 0.37$). Tukey HSD tests showed that for the form-signal-randomized stimuli, while the decoding accuracy for the motion-defined FoE shift was significantly higher than the baseline level for all the six visual areas ($p < 0.00039$), the decoding accuracy for the form-defined FoE shift was not statistically different from the baseline level for all the six visual areas ($p > 0.91$). In contrast, for the motion-signal-randomized stimuli, while the decoding accuracy for the form-defined FoE shift was significantly higher than the baseline level for all the six visual areas ($p < 0.011$), the decoding accuracy for the motion-defined FoE shift was not statistically different from the baseline level for all the six visual areas ($p > 0.68$). These results show that randomizing form signals to remove the form cue to the FoE indeed only affected the decoding accuracy for the form-defined FoE shift, and randomizing motion signals to remove the motion cue to the FoE indeed only affected the decoding accuracy for the motion-defined FoE shift. This supports the claim that the responses in the brain areas identified in Experiments 1 and 2 are driven by global form and motion signals.

Because randomizing form or motion signals removed the form or the motion cue to the FoE in the four stimuli of Experiment 2, it made the two congruent stimuli the same as the two incongruent stimuli. We thus expected that all the six visual areas would show similar decoding accuracies for the congruent and the incongruent stimuli. To examine this, as in Experiment 2, we trained a two-way linear SVM classifier to discriminate the pattern of BOLD response to the two congruent or incongruent stimuli using the data from three pairs of even- and odd-numbered runs randomly selected from the eight experimental runs. The classifier's prediction accuracy of the two congruent or incongruent stimuli was calculated using the data from the remaining pair of even- and odd-numbered experimental runs. This procedure was repeated 16 times to cover all possible combinations of even- and odd-numbered runs for training and testing. Figure 6*c* plots the classifier's mean prediction accuracy (i.e., decoding accuracy) of the two congruent or incongruent stimuli averaged across 16 repetitions for the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right) for the six visual areas. A 2 (decoding vs baseline decoding accuracy) × 2 (congruent vs incongruent stimuli) repeated-measures ANOVA revealed that, for both the form- and the motion-signal-randomized stimuli, only the main effect of decoding accuracy was significant for all six visual areas ($F_{(1,11)} > 8.95$, $p < 0.013$, $\eta_p^2 > 0.45$). That is, across the congruent and the incongruent stimuli, the decoding accuracy was significantly higher than the baseline decoding accuracy for all six visual areas. Tukey HSD tests showed that, for both the form- and the motion-signal-randomized stimuli, there was no significant difference in the classifier's decoding accuracy between the congruent and the incongruent stimuli for all six visual areas ($p > 0.12$). This confirms that after randomizing form or motion signals to render the two congruent stimuli the same as the two incongruent stimuli, all the six visual areas indeed could not tell the difference between the congruent and the incongruent stimuli anymore.

Eye movement data

The recorded eye movement data are plotted in Figure 7. For Experiment 1, a one-way repeated-measures ANOVA (with the Greenhouse–Geisser correction for any lack of sphericity) revealed no significant difference in the horizontal ($F_{(5,25)} = 1.54$, $p = 0.21$, $\eta_p^2 = 0.24$) or vertical ($F_{(2.5,12.5)} = 0.28$, $p = 0.81$, $\eta_p^2 = 0.053$) eye positions across the six stimuli. There was also no significant difference in saccade amplitude ($F_{(5,25)} = 0.39$, $p = 0.85$, $\eta_p^2 = 0.072$) or the number of saccades ($F_{(1.55,7.75)} = 0.997$, $p = 0.39$, $\eta_p^2 = 0.17$) across the six stimuli. For Experiment 2, a one-way repeated-measures ANOVA (with the Greenhouse–Geisser correction for any lack of sphericity) revealed no significant difference in the horizontal ($F_{(1.01,5.07)} = 2.69$, $p = 0.16$, $\eta_p^2 = 0.35$) or vertical ($F_{(3,15)} = 0.85$, $p = 0.49$, $\eta_p^2 = 0.15$) eye positions across the four stimuli. There was also no significance difference in saccade amplitude ($F_{(3,15)} = 1.95$, $p = 0.17$, $\eta_p^2 = 0.28$) or the number of saccades ($F_{(3,15)} = 1.59$, $p = 0.23$, $\eta_p^2 = 0.24$) across the four stimuli. These results support the claim that participants were able to follow our instructions and adequately maintain their eye position at the center of the display throughout the trial.

Discussion

Combining the results from the three experiments, we found that the early visual areas V1, V2, and V3 (V3v and V3d combined) respond to a position shift of the FoE defined by either motion or form cues. This is consistent with the findings of primate neurophysiology studies showing that these areas process both local motion and form information (Hubel and Wiesel, 1968; Mikami et al., 1986; Felleman and Van Essen, 1987; Levitt et al., 1994; Gegenfurtner et al., 1997; Hu et al., 2018). Research identifying

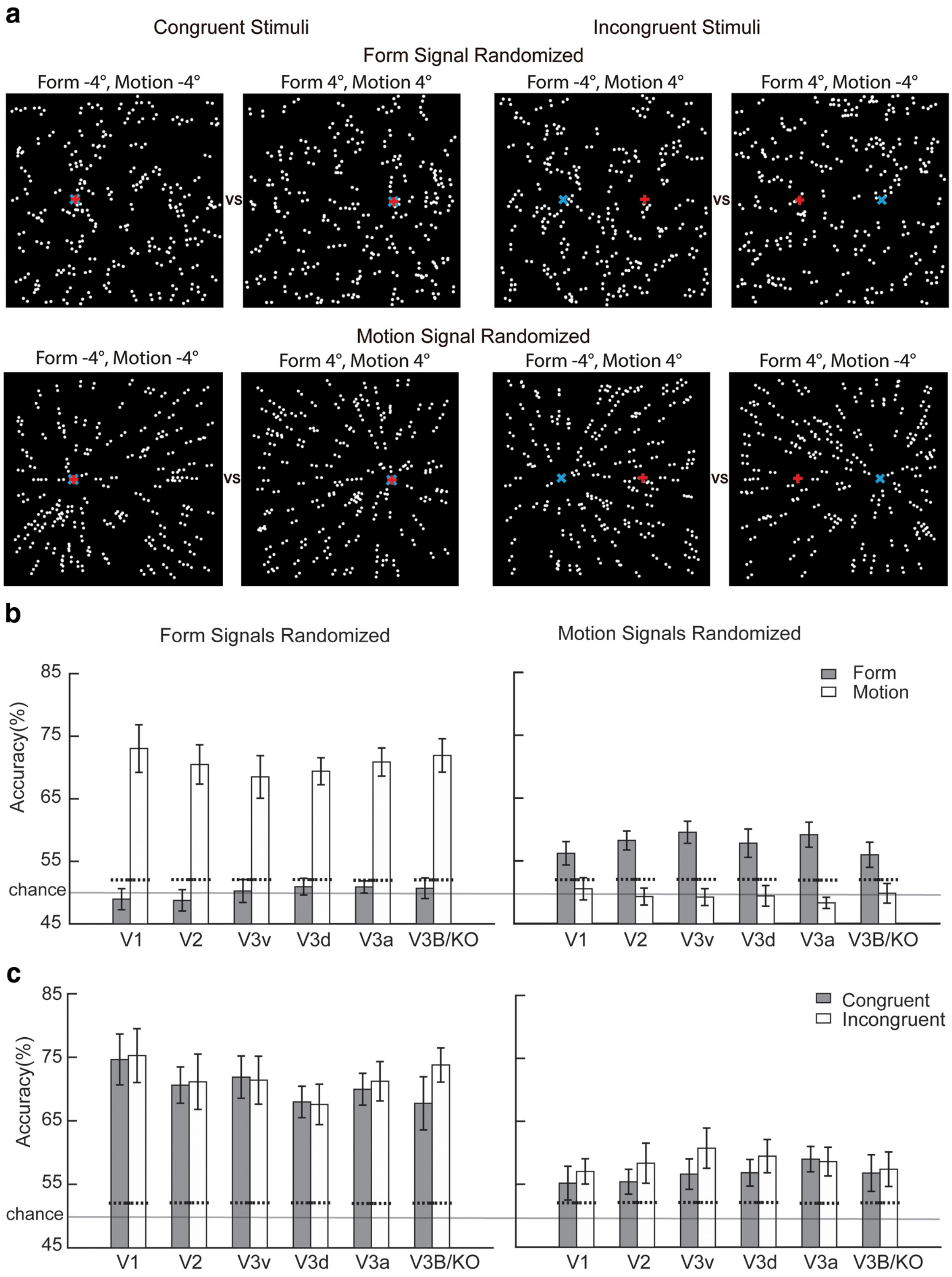


Figure 6. Experiment 3 visual stimuli and data. ***a***, Illustrations of the stimuli with the form signals randomized and the motion signals intact (top row) and the stimuli with the motion signals randomized and the form signals intact (bottom row). Negative sign indicates the FoE location to the left of the display center; positive sign indicates the FoE location to the right. The "×" and the "+" indicate the form- and the motion-defined FoEs, respectively. ***b***, The classifier's decoding accuracy for the form- (gray) or the motion-defined FoE shift (white) with the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right) for the six visual areas. ***c***, The classifier's decoding accuracy for the congruent (gray) or incongruent (white) stimuli with the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right) for the six visual areas. Dotted lines indicate the upper limits of the 95% CI of the classifier's baseline decoding accuracies. Solid line indicates the chance level of 1/2. Error bars indicate SEs across 12 participants.

the homology of primate areas V1, V2, and V3 in the human brain has been quite successful and shows that these areas in humans are organizationally and functionally analogous to those in macaques. However, for visual areas beyond V3, the homology between the primate and human brain breaks down and is less certain (Winawer and Witthoft, 2015).

Our results show that after area V3, the dorsal (V3a and V3B/KO) rather than the ventral visual areas (hV4 and LO) respond to

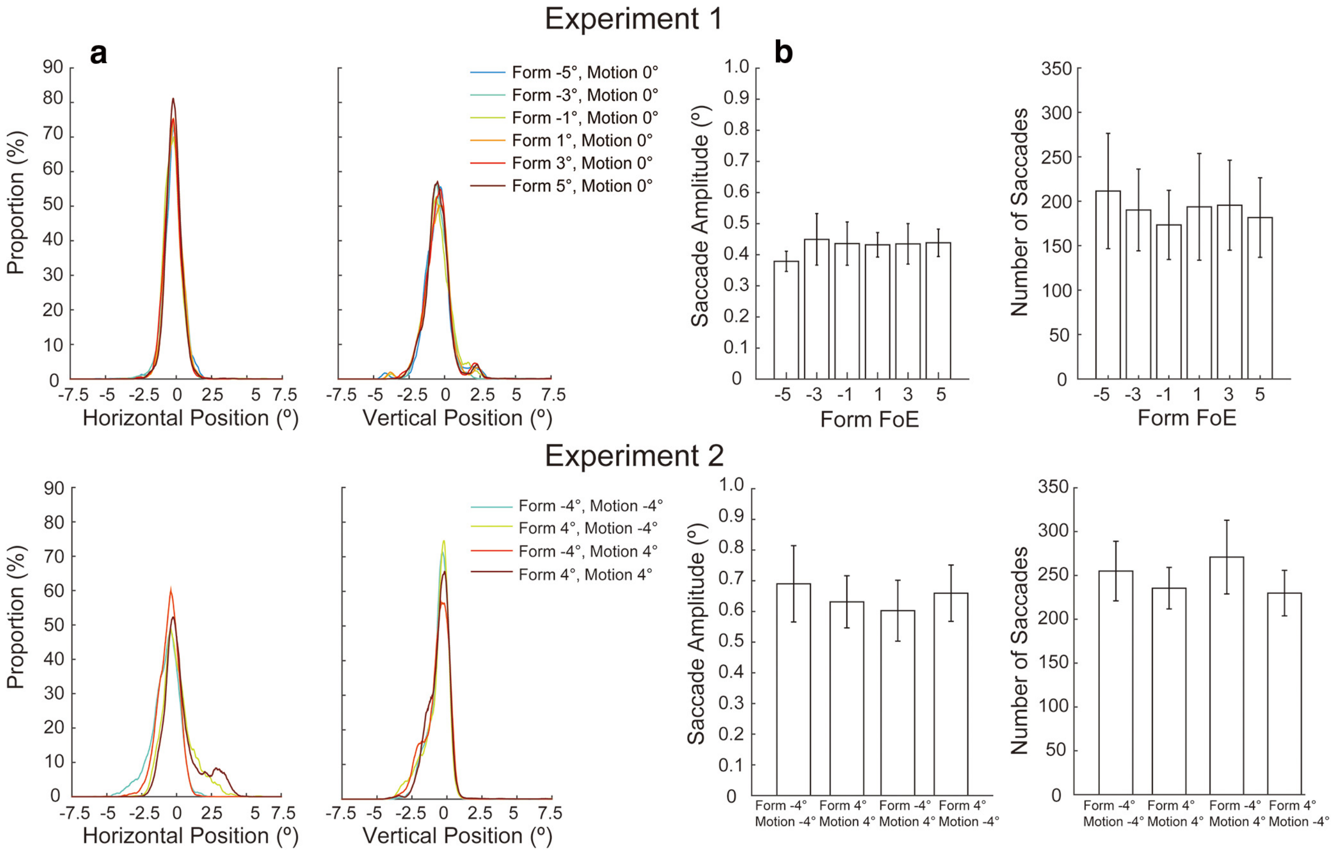


Figure 7. Eye movement data. ***a***, The proportion of eye position data of 6 participants as a function of the deviation between eye fixation and the center of the display along horizontal and vertical directions for the six stimuli in Experiment 1 (top) and the four stimuli in Experiment 2 (bottom). ***b***, The saccade amplitude and the number of saccades against the six stimuli in Experiment 1 (top) and the four stimuli in Experiment 2 (bottom). Error bars indicate SEs across 6 participants.

either the motion- or the form-defined FOE shift. Previous research has shown that the center of a radial flow pattern activates area V3a, suggesting that this area responds to the exact location of the FoE in optic flow (Koyama et al., 2005). Our finding regarding area V3a thus complements previous findings for this area. Our findings are also consistent with the dissociation of the ventral and dorsal streams regarding visual information processing for perception and action (Goodale and Milner, 1992). Specifically, the ventral stream recognizes and discriminates shape, size, and color of objects (Kravitz et al., 2013) and thus supports vision for perception, whereas the dorsal stream encodes spatial location, orientation, and motion of objects to guide actions and thus supports vision for action (Decety and Grèzes, 1999). Because our stimuli provide heading information that can be used for the control of self-motion (e.g., Gibson, 1950; L. Li and Niehorster, 2014), it is reasonable that the dorsal but not ventral visual areas respond to the motion- or the form-defined FoE shift.

The data of Experiment 1 show that, after area V3B/KO, no other high-level brain areas appear to respond to the form-defined FoE shift. The data of Experiment 2 further show that area V3B/KO shows a highly significant higher decoding accuracy for the congruent than the incongruent stimuli, and its decoding sensitivity to the combined motion and form cues is higher than the quadratic sum of its decoding sensitivity to each cue alone. This suggests that area V3B/KO does not perform a simple linear summation of motion and form information but fuses or integrates these two types of information to form a unified representation or percept. This is consistent with anatomical and function roles of area V3B/KO in visual information processing. Anatomically, humanV3B/KO corresponds to the dorsal portion of primate V4 that receives inputs from the earlier visual area (Zeki, 2003). More recent studies identified that the dorsal endpoints of the vertical occipital fasciculus, the only major fiber bundle connecting occipital dorsal and ventral streams (Yeatman et al., 2014), are near area V3B/KO and its neighboring area, such as area V3d (Takemura et al., 2016). Functionally, V3B/KO is originally defined as the kinetic occipital area that responds to shapes generated from kinetic boundaries (DuPont et al., 1997) and implied motion (Krekelberg et al., 2005). Several brain imaging studies also provide evidence for the involvement of area V3B/KO in processing optic flow (Greenlee, 2000; Rutschmann et al., 2000; Beer et al., 2002) and global form structure (S. Li et al., 2007; Ostwald et al., 2008). Area V3B/KO could thus naturally integrate form and motion signals when they are both available for the perception of heading.

As an area of cue integration, V3B/KO should deal with conflicting signals and decide whether or not to combine the cues. Using single-neuron recording in macaque monkeys, Gu et al. (2008) have shown that area MSTd, which integrates visual and vestibular cues, contains neurons that are best stimulated by a discrepancy between these cues. Rideaux and Welchman (2018) developed a model based on their data and proposed that such neurons also exist in the human brain, such as in area V3B/KO, to provide "what not" information that drives suppression of integration when the discrepancy is large. It is possible that the early visual areas that encode the discrepancy between motion- and form-defined FoEs feed into area V3B/KO for its population of "what not" neurons to decide when to combine motion and form cues. The output of V3B/KO may have similar responses to stimuli that can be integrated; thus, its response is not modulated by the magnitude of the position shift in the form-defined FoE (see Figure 1*d*).

Previous studies have also shown that area V3B/KO is a candidate cortical locus for the integration of qualitatively different cues. For example, Ban et al. (2012) found that area V3B/KO integrates disparity and motion information for depth perception. It has been further revealed that the cue integration in area V3B/KO is not specific to particular cue pairing, such as disparity and motion, but can be generalized to other cue pairings, such as disparity and shading (Dövencioglu et al., 2013) or disparity and texture (Murphy et al., 2013). Using transcranial direct current stimulation to perturb the excitatory and inhibitory balance of areaV3B/KO led to impaired performance of such cue integration (Rideaux and Welchman, 2018). Our study used quite different types of cues from those examined by previous studies, i.e., the motion- and form-defined FoEs that are also qualitatively different. The finding that these cues are integrated in area V3B/KO for the perception of heading is thus compatible with previous findings and suggests quite general cue integration computations in area V3B/KO that could not be inferred from previous studies.

The results of the current study do not provide evidence for the involvement of the dorsal motion (MT and MST) or other optic flow visual areas (VIP, V6, and CSv) in the integration of motion and form cues for the perception of heading. We do not exclude the possibility that this could be due to the sampling and measurement approach we took in the current study. For example, we localized area VIP primarily based on its anatomical structure described in previous studies. Given the variation in peak locations between different studies and the variations between participants, our localization of area VIP might not be precise. Nevertheless, the searchlight analysis results confirm that this area did not respond to the form-defined FoE shift. We believe that the lack of response to the form-defined FoE shift in the dorsal motion or other optic flow areas is more related to the ability to encode fine differences in the FoE location using the activity of spatially precise receptive fields in early visual areas. For example, studies have shown that the human homolog of primate MST can discriminate expansion from contraction flow patterns but does not appear to encode the specific location of the FoE in optic flow (Strong et al., 2017), and human V6 is also not sensitive to the change in location of the FoE in optic flow (Furlan et al., 2014). In addition, previous findings of primate neurophysiology studies have shown that most MST neurons do not respond to form information (Geesaman and Andersen, 1996), and area VIP receives a large amount of input from area MST but not much input from the ventral stream (Ungerleider et al., 2008). In contrast to other flow-selective brain areas, CSv responses to optic flow can be suppressed by many factors, such as whether the flow pattern is compatible with self-motion (Wall and Smith, 2008) or whether flow is used for visuomotor control (Field et al., 2015). All these factors can contribute to the lack of response in higher visual areas associated with optic flow processing to the form-defined FoE shift and thus the lack of involvement in the integration of motion and form cues for the perception of heading.

In conclusion, using the fMRI and MVPA technique, our study systematically examined human brain areas that integrate motion and form cues for the perception of the direction of self-motion (i.e., heading). Our results show that motion and form cues are first processed in the early visual areas and then are likely integrated in the higher dorsal area V3B/KO for the final estimation of heading during self-motion.

References

Ban H, Preston TJ, Meeson A, Welchman AE (2012) The integration of motion and disparity cues to depth in dorsal visual cortex. Nat Neurosci 15:636–643.

Beer J, Blakemore C, Previc FH, Liotti M (2002) Areas of the human brain activated by ambient visual motion, indicating three kinds of self-movement. Exp Brain Res 143:78–88.

Benson DF, Greenberg JP (1969) Visual form agnosia: aspecific defect in visual discrimination. Arch Neurol 20:82–89.

Braddick OJ, O'Brien JM, Wattam-Bell J, Atkinson J, Turner R (2000) Form and motion coherence activate independent, but not dorsal/ventral segregated, networks in the human brain. Curr Biol 10:731–734.

Burr DC, Ross J (2002) Direct evidence that "speedlines" influence motion mechanisms. J Neurosci 22:8661–8664.

Crowell JA, Banks MS (1993) Perceiving heading with different retinal regions and types of optic flow. Percept Psychophys 53:325–337.

Decety J, Grèzes J (1999) Neural mechanisms subserving the perception of human actions. Trends Cogn Sci 3:172–178.

DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. Trends Neurosci 11:219–226.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci U S A 93:2382–2386.

Dövencioglu D, Ban H, Schofield AJ, Welchman AE (2013) Perceptual integration for qualitatively different 3-D cues in the human brain. J Cogn Neurosci 25:1527–1541.

Dukelow SP, DeSouza JF, Culham JC, van den Berg AV, Menon RS, Vilis T (2001) Distinguishing subregions of the human MT plus complex using visual fields and pursuit eye movements. J Neurophysiol 86:1991–2000.

Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marchal G, Mortelmans L, Orban GA (1997) The kinetic occipital region in human visual cortex. Cereb Cortex 7:283–292.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.

Felleman DJ, Van Essen DC (1987) Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. J Neurophysiol 57:889–920.

Field DT, Inman LA, Li L (2015) Visual processing of optic flow and motor control in the human posterior cingulate sulcus. Cortex 71:377–389.

Furlan M, Wann JP, Smith AT (2014) A representation of changing heading direction in human cortical areas pVIP and CSv. Cereb Cortex 24:2848–2858.

Geesaman BJ, Andersen RA (1996) The analysis of complex motion patterns by form/cue invariant MSTd neurons. J Neurosci 16:4716–4732.

Gegenfurtner KR, Kiper DC, Levitt JB (1997) Functional properties of neurons in macaque area V3. J Neurophysiol 77:1906–1923.

Geisler WS (1999) Motion streaks provide a spatial code for motion direction. Nature 400:65–69.

Gibson JJ (1950) The perception of the visual world. Boston: Houghton Mifflin.

Glass L (1969) Moire effect from random dots. Nature 223:578–580.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20–25.

Greenlee MW (2000) Human cortical areas underlying the perception of optic flow: brain imaging studies. Int Rev Neurobiol 44:269–292.

Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201–1210.

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Hu J, Ma H, Zhu S, Li P, Xu H, Fang Y, Chen M, Han C, Fang C, Cai X, Yan K, Lu HD (2018) Visual motion processing in macaque V2. Cell Rep 25:157–167.e5.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215–243.

Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept Psychophys 14:201–211.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. Science 293:1506–1509.

Kourtzi Z, Krekelberg B, van Wezel RJ (2008) Linking form and motion in the primate brain. Trends Cogn Sci 12:230–236.

Koyama S, Sasaki Y, Andersen GJ, Tootell RB, Matsuura M, Watanabe T (2005) Separate processing of different global-motion structures in visual cortex is revealed by FMRI. Curr Biol 15:2027–2032.

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn Sci 17:26–49.

Krekelberg B, Vatakis A, Kourtzi Z (2005) Implied motion from form in the human visual cortex. J Neurophysiol 94:4373–4386.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Res 35:389–412.

Levitt JB, Kiper DC, Movshon JA (1994) Receptive fields and functional architecture of macaque V2. J Neurophysiol 71:2517–2542.

Li L, Niehorster DC (2014) Influence of optic flow on the control of heading and target egocentric direction during steering toward a goal. J Neurophysiol 112:766–777.

Li L, Peli E, Warren WH (2002) Heading perception in patients with advanced retinitis pigmentosa. Optom Vis Sci 79:581–589.

Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain. J Neurosci 27:12321–12330.

Mikami A, Newsome WT, Wurtz RH (1986) Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. J Neurophysiol 55:1328–1339.

Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414–417.

Murphy AP, Ban H, Welchman AE (2013) Integration of texture and disparity cues to surface slant in dorsal visual cortex. J Neurophysiol 110:190–203.

Niehorster DC, Cheng JC, Li L (2010) Optimal combination of form and motion cues in human heading perception. J Vis 10:20.

Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn Sci 8:315–324.

Orban GA, Claeys K, Nelissen K, Smans R, Sunaert S, Todd JT, Wardak C, Durand JB, Vanduffel W (2006) Mapping the parietal cortex of human and non-human primates. Neuropsychologia 44:2647–2667.

Ostwald D, Lam JM, Li S, Kourtzi Z (2008) Neural coding of global form in the human visual cortex. J Neurophysiol 99:2456–2469.

Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Patria F, Galletti C (2010) Human V6: the medial motion area. Cereb Cortex 20:411–424.

Rideaux R, Welchman AE (2018) Proscription supports robust perceptual integration by suppression in human visual cortex. Nat Commun 9:1502.

Rutschmann RM, Schrauf M, Greenlee MW (2000) Brain activation during dichoptic presentation of optic flow stimuli. Exp Brain Res 134:533–537.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Strong SL, Silson EH, Gouws AD, Morland AB, McKeefry DJ (2017) A direct demonstration of functional differences between subdivisions of human V5/MT+. Cereb Cortex 27:1–10.

Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, Pestilli F (2016) A major human white matter pathway between dorsal and ventral visual cortex. Cereb Cortex 26:2205–2214.

Ungerleider LG, Galkin TW, Desimone R, Gattass R (2008) Cortical connections of area V4 in the macaque. Cereb Cortex 18:477–499.

van den Berg AV (1992) Robustness of perception of heading from optic flow. Vision Res 32:1285–1296.

Wallach H, O'Connell DN (1953) The kinetic depth effect. J Exp Psychol 45:205–217.

Wall MB, Smith AT (2008) The representation of egomotion in the human brain. Curr Biol 18:191–194.

Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. Neuron 56:366–383.

Warren WH Jr, Morris MW, Kalish M (1988) Perception of translational heading from optical flow. J Exp Psychol Hum Percept Perform 14:646–660.

Winawer J, Witthoft N (2015) Human V4 and ventral occipital retinotopic maps. Vis Neurosci 32:E020.

Yeatman JD, Weiner KS, Pestilli F, Rokem A, Mezer A, Wandell BA (2014) The vertical occipital fasciculus: a century of controversy resolved by in vivo measurements. Proc Natl Acad Sci U S A 111:E5214–E5223.

Zeki S (2003) Improbable areas in the visual brain. Trends Neurosci 26:23–26.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11:641–649.

Zeki S, Perry RJ, Bartels A (2003) The processing of kinetic contours in the brain. Cereb Cortex 13:189–202.

Zihl J, von Cramon D, Mai N (1983) Selective disturbance of movement vision after bilateral brain damage. Brain 106:313–340.