

Integration of motion and form cues for the perception of self-motion in the human brain

Article

Accepted Version

Kuai, S.-G., Shan, Z.-K.-D., Chen, J., Xu, Z.-X., Li, J.-M., Field, D. T. ORCID: <https://orcid.org/0000-0003-4041-8404> and Li, L. (2020) Integration of motion and form cues for the perception of self-motion in the human brain. *The Journal of Neuroscience*, 40 (5). pp. 1120-1132. ISSN 1529-2401 doi: <https://doi.org/10.1523/JNEUROSCI.3225-18.2019> Available at <https://centaur.reading.ac.uk/88279/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1523/JNEUROSCI.3225-18.2019>

Publisher: The Society for Neuroscience

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

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^{3,2}, Jing Chen^{3,2}, Zhe-Xin Xu¹, Jia-Mei Li¹, Qi Liang¹, David T Field⁴, Li Li^{3,2*}

¹ Shanghai Key Laboratory of Brain Functional Genomics, Key Laboratory of Brain Functional Genomics, Ministry of Education, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China.

² NYU-ECNU Institute of Brain and Cognitive Science at New York University Shanghai, Shanghai, China.

³ Faculty of Arts and Science, New York University Shanghai, Shanghai, China.

⁴ Center for Integrative Neuroscience & Neurodynamics, Department of Psychology, University of Reading, UK.

Corresponding authors:

Li Li
NYU Shanghai
1555 Century Avenue
Pudong New Area
Shanghai
200122
PRC
LL114@nyu.edu

Shuang Kuai
East China Normal University
3663 Zhongshan Road North
Putuo
Shanghai
200062
PRC
sgkuai@psy.ecnu.edu.cn

Abbreviated title: Motion and form cues for heading perception

Pages: 35

Figures: 6; tables: 0; multimedia and 3D models: 1; Supplemental figures: 2

Abstract: 244 words; Introduction: 649 words; Discussion: 1441 words

Note: Figures embedded (with captions) to enhance readability for manuscript review

Acknowledgements

This study was supported by research grants from the National Natural Science Foundation of China (31741061, 31771209, and 31571160), the National Social Science Foundation of China (15ZDB016), Shanghai Science and Technology Committee (15DZ2270400, 17ZR1420100), China Ministry of Education (ECNU 111 Project, Base B1601), and NYU-ECNU Joint Research Institute at NYU Shanghai. SGK and LL designed the experiments. ZKDS, ZXX, JML, and JC ran the experiments. All analysed the data. The paper was written by LL, SGK, and JC. The authors declare no competing financial interests.

1 **Abstract (244/250)**

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

18 **Significance statement (120/120)**

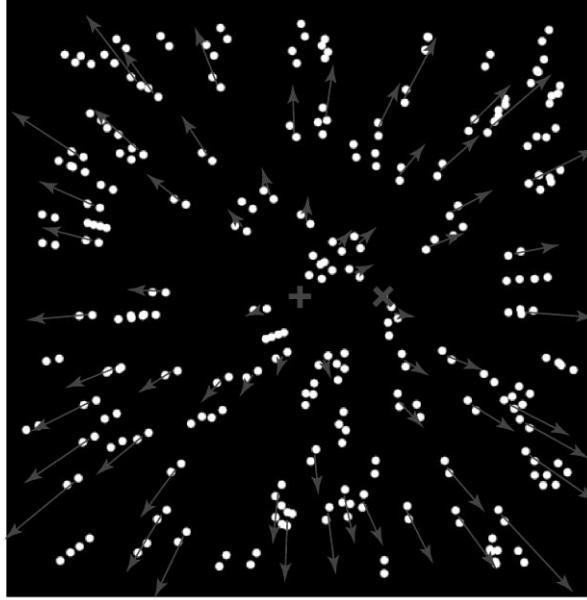
19 Human survival relies on accurate perception of self-motion. The visual system uses both motion (optic
20 flow) and form cues for the perception of the direction of self-motion (heading). Although human brain
21 areas for processing optic flow and form structure are well identified, the areas responsible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

27 **Introduction (649/650)**

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

34 Although the FoE is defined by the expanding global motion in optic flow, it is also given by global
35 form information such as motion streaks in a time-integrated flow field. Since Gibson's proposal, research
36 has focused almost exclusively on what motion cues people use to perceive heading but ignored the
37 potential influence of form cues. This could be partly due to the proposal (e.g., Mishkin et al., 1983;
38 DeYoe and Van Essen, 1988) that motion and form cues are processed with two separate visual streams
39 that originate from the primary visual cortex and project either dorsally to the parietal cortex for motion
40 processing or ventrally to the inferotemporal cortex for form processing.

41 Separate processing of motion and form information is initially supported by neuropsychological
42 evidence from brain-damaged patients (e.g., Benson and Greenberg, 1969; Zihl et al., 1983; Goodale and
43 Milner, 1992). However, many studies show that motion and form processing are closely linked (see
44 Kourtzi et al., 2008 for a review). For example, the classical kinetic depth effect (Wallach and O'Connell,
45 1953) and biological motion (Johansson, 1973) show that motion can help perceive form that could not be
46 seen from a static display. Conversely, form can also affect motion perception – static “speed lines”
47 (motion streaks) depicted in cartoons are shown to bias the perceived object motion direction (e.g.,
48 Geisler, 1999; Burr and Ross, 2002).



49

50 Figure 1. Illustrations of an animated Glass pattern stimulus that offers two
51 independent FoEs: the form-defined FoE given by the orientation of the dot pairs (“x”)
52 and the motion-defined FoE given by the motion of the dot pairs (“+”). Lines with
53 arrowheads represent velocity vectors of the centroid of the dot pairs. “x”, “+”, and
54 lines with arrow heads are for illustration purpose only and not shown in the
55 experimental stimulus.

56 Enlightened by these studies, Niehorster et al. (2010) developed animated Glass pattern stimuli
57 (Glass, 1969) that pitted optic flow and form cues to self-motion against one another with each cue
58 indicating a different heading direction. They for the first time found that the human visual system
59 optimally integrates flow and form cues for heading estimation. Although the brain areas for processing
60 flow and form cues are well identified, the areas responsible for integrating these two cues for the
61 perception of self-motion remain unknown. To address this question, in the current study, we used similar
62 animated Glass pattern stimuli consisting of randomly distributed dot pairs oriented toward a locus on a
63 screen (the form-defined FoE) but moved away from a different locus (the motion-defined FoE) to
64 simulate observer translation (see Figure 1 and [Movie 1](#)). In Experiment 1, we fixed the motion-defined
65 FoE location and shifted the form-defined FoE location. In Experiment 2, we made the locations of the
66 motion- and the form-defined FoEs either congruent (at the same location in the display) or incongruent
67 (on the opposite sides of the display). The shift in location of the motion- or the form-defined FoE was
68 the same in the two types of stimuli but the perceived direction of heading shifted for the congruent but

69 not the incongruent stimuli. We performed searchlight and ROI-based multiple voxel pattern analysis
70 (MVPA) to find the brain areas that could not only respond to a location shift of the form-defined FoE
71 (Experiment 1) but also show a higher decoding accuracy for the congruent than the incongruent stimuli
72 (Experiment 2). These areas are likely to be in charge of integrating motion and form cues for heading
73 perception. In Experiment 3, we randomized the form or the motion signals in the stimuli to remove the
74 form or the motion cues to the FoE. The purpose was to validate whether the cortical areas identified in
75 Experiments 1 and 2 are indeed driven by global form and motion signals.

76 **Materials and Methods**

77 **Experimental Design and Statistical Analyses**

78 The experiments were within-subject designs. Data were analyzed using repeated-measures ANOVAs
79 and *t*-tests. We reported exact *p* values except when $p < 0.001$. We report η^2 and Cohen's *d* as a measure
80 of effect size for ANOVAs and *t*-tests, respectively.

81 **Participants**

82 Twenty-six students and staff (22 naïve to the specific goals of the study) between the age of 18 and
83 38 at East China Normal University (ECNU) and New York University Shanghai (NYU SH) participated
84 in the study. Among them, 14 (9 males, 5 females; mean age \pm SD: 23.4 ± 5.92) participated in
85 Experiment 1, 13 (9 males, 4 females; mean age \pm SD: 22.8 ± 4.28) participated in Experiment 2, and 12
86 (5 males, 7 females; mean age \pm SD: 23.5 ± 2.15) participated in Experiment 3. Participants of
87 Experiments 1 and 2 also participated in a control psychophysical experiment.

88 All participants had normal or corrected-to-normal vision and provided informed consent. The study
89 was approved by the Human Research Ethics Committee at ECNU and the Internal Review Board at
90 NYU SH. We determined the sample size based on the sample size in relevant previous studies.

91 **Visual stimuli**

92 The display simulated an observer translating at 1.5 m/s through a 3D cloud consisting of 200 white
93 dot pairs with 0.25° centroid-to-centroid separation (dots: 0.125° in diameter, 95% luminance contrast).
94 The 200 dot pairs were randomly placed in the depth range of 1.1–5 m such that the same number of dot
95 pairs originated from each distance in depth. Dot pairs moved outside of the field of view were
96 regenerated with an algorithm that maintained the depth layout of the 3D cloud. In each frame, all dot
97 pairs were oriented toward a location on the screen forming a radial Glass pattern. The display thus
98 offered two independently generated FoEs: the form-defined FoE given by the orientation of the dot pairs
99 (“×” in Figure 1) and the motion-defined FoE given by the centroid of dot pairs moved outward (“+” in
100 Figure 1).

101 In Experiment 1, the motion-defined FoE was fixed at 0° (the center of the display) and the form-
102 defined FoE was shifted from -5° (left) to 5° (right) in steps of 2° from the motion-defined FoE, resulting
103 in six stimuli (Figure 2a). In Experiment 2, we tested two congruent and two incongruent stimuli. For the
104 two congruent stimuli, the motion- and the form-defined FoEs were both at -4° or 4° . For the two
105 incongruent stimuli, the form- and the motion-defined FoEs were at 4° on the opposite sides of the
106 display (Figure 3a). In Experiment 3, we used the four stimuli in Experiment 2 and randomized the
107 orientation of the dot pairs or the motion direction of the dot pairs, resulting in four form-signal-
108 randomized stimuli and four motion-signal-randomized stimuli. Randomizing the orientation of the dot
109 pairs removed the form-defined FoE but left the motion-defined FoE intact (Figure 6a, top row), and
110 randomizing the motion direction of the dot pairs removed the motion-defined FoE but left the form-
111 defined FoE intact (Figure 6a, bottom row).

112 On each trial, a red fixation point appeared at the center of the display for 400 ms followed by the
113 simulated self-motion display for 600 ms. No fixation point was present in the self-motion display to
114 ensure that the self-motion display did not contain any extraneous relative motion. Participants were

115 instructed to fixate the fixation point that appeared at the beginning of the trial and maintain their eye
116 position at the center of the display throughout the trial. If participants followed our instructions, then the
117 pattern of their eye movements should not vary across the stimulus conditions in all experiments. In 20%
118 of trials, the contrast of half of the dot pairs was lowered by about 50%. Participants were asked to watch
119 the display carefully and press a button to report the trials containing dots with lower contrast.

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

138 **Equipment and imaging acquisition parameters**

139 The display was rendered with Psychtoolbox-3 Toolbox and back projected on a white screen
140 (resolution: 1024H × 768V pixels; refresh rate: 60 Hz) in a Siemens Magnetom Prisma 3T MRI scanner.
141 Participants lay supine in the scanner and viewed the display (19°×19°) binocularly through light
142 reflecting mirrors **at the distance of 92 cm**. Participants' head was positioned in a 32-channel head coil for
143 enhanced signal-to-noise. Functional scans consisted of repeated echo-planar imaging (EPI): voxel size =
144 3 × 3 × 4 mm¹, echo time (TE) = 30 ms, flip angle (FA) = 81°, matrix size = 64 × 64, field of view (FOV)
145 = 192 × 192 mm², with slice order ascending and interleaved, 38 slices (inter-slice gap = 0.3 mm, slice
146 thickness = 3.0 mm), and repetition time (TR) = 2000 ms. A detailed T1-weighted anatomical image was
147 acquired (voxel size = 1 × 1 × 1 mm, TE = 2.34 ms, FA = 7°, FOV = 256 × 256 mm², 192 slices, no gap,
148 TR = 2530 ms, total scan time = 5 min and 48 s).

149 In the psychophysical experiment, the display was presented on an ASUS VG278H 27-inch LCD
150 monitor (resolution: 1024H × 768V pixels; refresh rate: 60 Hz). Participants viewed the display (19° ×
151 19°) binocularly with their head stabilized by a chin rest **at 57 cm away from the display**.

152 **Procedure**

153 In all three experiments, participants were scanned for eight runs using a block design. Each run had
154 24 stimulus blocks (6 stimuli × 4 blocks) in Experiment 1, 16 stimulus blocks (4 stimuli × 4 blocks) in
155 Experiment 2, and **24 stimulus blocks (8 stimuli × 3 blocks) in Experiment 3**. Each 16-s stimulus block
156 contained 16 trials of a stimulus. The testing order of stimulus was randomized in each run. Each run also
157 had a 16-s fixation block with no stimulus but a red fixation point in the center of a blank screen at the
158 beginning, in the middle, and at the end of the run. The purpose of the fixation block was to acquire
159 baseline brain activations in each run. The scanning lasted about 1 hr for Experiment 1, about 40 min for
160 Experiment 2, **and about 1 hr for Experiment 3**.

¹ Voxel size was 3 × 3 × 3 mm in Experiment 3 due to a system upgrade.

161 For each participant in a separate scanning session that lasted about 1 hr, we identified the following
162 regions of interest (ROI): the early visual areas that respond to both local motion and form information
163 (V1, V2), the higher ventral areas that respond to shape and global form information (V3v, hV4, LO), the
164 dorsal (hMST) and the parietal areas (VIP, V6) and area CSv that respond to optic flow. Because previous
165 human brain-imaging studies have shown that the dorsal stream can be activated by both motion and form
166 information (Braddick et al., 2000; Krekelberg et al., 2005), we also identified other visual areas along the
167 dorsal stream (V3d, V3a, V7, V3B/KO, hMT) that are known to respond to motion information.
168 Specifically, we identified the retinotopic visual areas (V1, V2, V3v, V3d, V3a, hV4, V7) using standard
169 retinotopic mapping procedures with rotating wedge stimuli (Engel et al., 1994; Sereno et al., 1995;
170 DeYoe et al., 1996). Area hV4 was defined as the ventral but not the dorsal sub-region of V4 (Wandell et
171 al., 2007). We identified areas V3B/KO (Dupont et al., 1997; Zeki et al., 2003), LO (Kourtzi and
172 Kanwisher, 2001), hMT (Zeki et al., 1991), hMST (Dukelow et al., 2001), V6 (Pitzalis et al., 2010), and
173 CSv (Wall and Smith, 2008) using independent localizers as described in the cited studies. Finally, we
174 identified area VIP (average center of ROI: -26, -64, 42 (left) and 28, -62, 47 (right); average number of
175 voxels: 255) by comparing the anatomical structure of the activated areas in the experiments to what is
176 described in previous studies (e.g., Orban et al., 2004; Orban et al., 2006).

177 To examine whether participants could follow our instructions to fixate the fixation point that
178 appeared at the center of the display at the beginning of a trial and then maintain their eye position there
179 throughout the trial, in a separated session outside of the scanner, we recorded eye movements of six
180 participants who all participated in Experiments 1 and 2 using an Eyelink 1000 plus eye tracker (1k Hz,
181 SR Research Ltd., Ontario, Canada) when they viewed the same display ($19^\circ \times 19^\circ$) on an LCD monitor
182 (1024×768 pixels, 60 Hz) and performed the same task as in Experiments 1 and 2.

183 In the psychophysical experiment, each participant completed a total of 260 experimental trials (4
184 stimulus conditions \times 13 FoE combinations \times 5 trials). The trials were blocked by stimulus condition and
185 randomized within each block. The testing order of stimulus condition was counterbalanced between

186 participants. Participants received 5-10 practice trials at the beginning of each block. No feedback was
187 provided in the practice or experimental trials. The psychophysical experiment lasted about 30 min.

188 **Data analysis**

189 *Pre-analysis*

190 Neuroimaging data were analyzed using Brain Voyager QX (Brain Innovations, Maastricht,
191 Netherlands). The anatomical data were transformed into the standard Montreal Neurological Institute
192 (MNI) space and then inflated using BrainVoyager QX. Pre-processing of the functional data included
193 slice scan time correction, 3D motion correction, linear trend removal, and temporal high-pass filtering.
194 The echo-planar imaging (EPI) images were then aligned with the anatomical images and transformed
195 into the standard MNI space. All functional data were transformed into a 3-mm isovoxel volume time
196 course (VTC) data using the nearest neighbor algorithm without spatial smoothing.

197 *Multi-voxel Pattern Analysis (MVPA)*

198 We performed MVPA (Haynes and Rees, 2005; Kamitani and Tong, 2005) to decode blood oxygen
199 level dependent (BOLD) responses evoked by different stimuli. We first normalized the time course data
200 by computing the Z-scores of BOLD signals in each run to minimize the baseline difference across runs.
201 We shifted the time course data forward by 4 s to compensate for the hemodynamic response delay and
202 then averaged the data across trials in each stimulus block. For the ROI-based analysis, we conducted a
203 general linear model (GLM) analysis to **select a number of the most activated voxels** in each ROI by
204 comparing their responses in the stimulus blocks with their baseline responses in the fixation blocks. For
205 the searchlight analysis (Kriegeskorte et al., 2006), we defined a spherical aperture (radius: 9 mm) and
206 moved this aperture voxel by voxel across the gray-matter of each participant's brain where the responses
207 in the stimulus blocks were higher than in the fixation blocks. We then trained a linear support vector
208 machine (SVM) classifier to discriminate the selected voxels' BOLD responses to different stimuli using
209 the data from seven out eight runs in the experiment and computed the accuracy of the classifier's

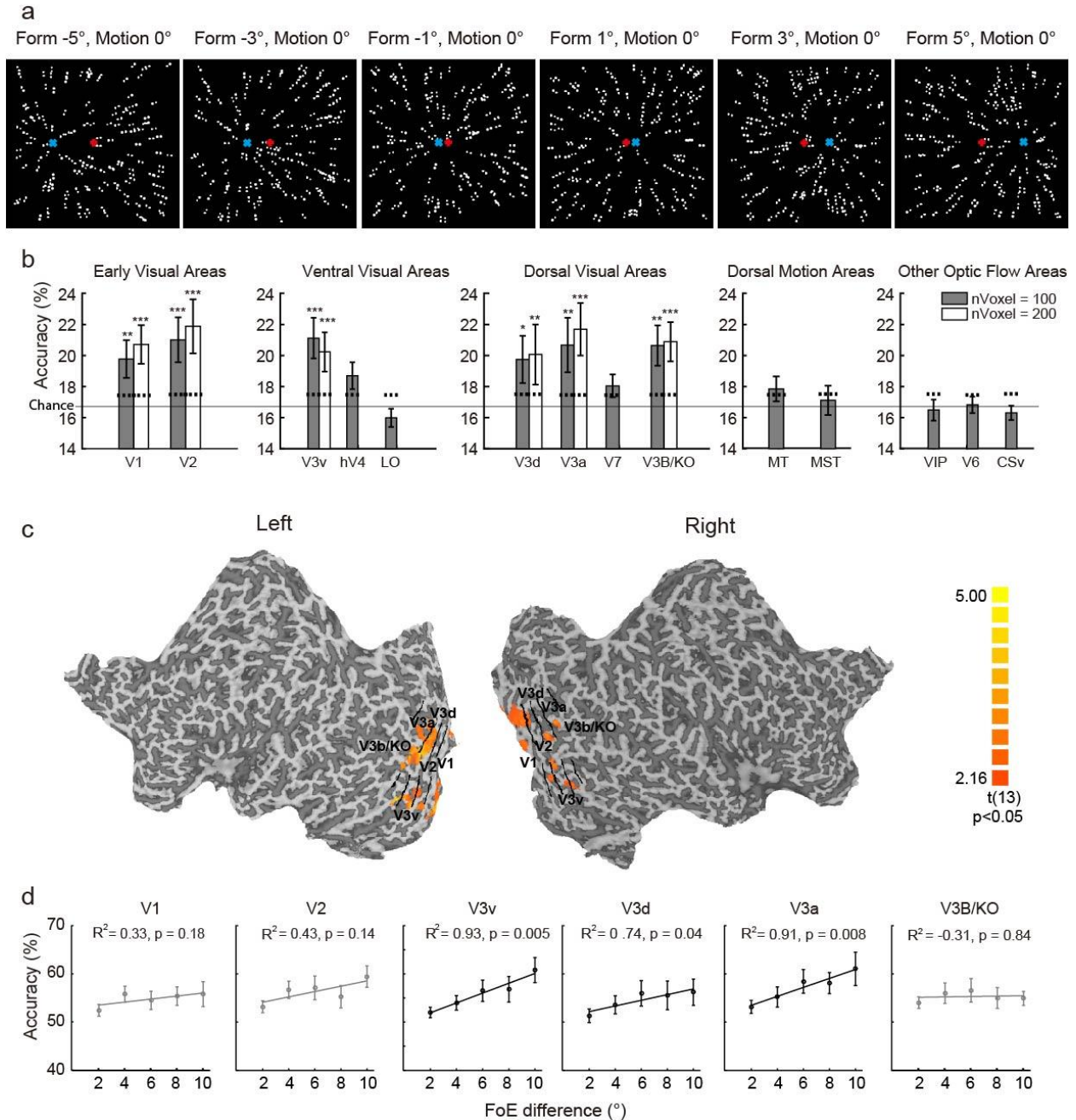
210 prediction of the stimuli in the unselected run. We repeated this procedure eight times to compute the
211 mean prediction accuracy averaged across eight runs, which was defined as the classifier's decoding
212 accuracy.

213 To estimate the significance level of the classifier's decoding accuracy, we performed a shuffled
214 analysis in which we randomly assigned the stimulus labels to the stimuli in the training stimulus blocks
215 and performed the same MVPA procedure for 1000 times. The computed mean prediction accuracy of the
216 stimuli in the testing stimulus blocks averaged across 1000 times was defined as the classifier's baseline
217 decoding accuracy.

218 **Results**

219 *Areas encoding form-defined FoEs*

220 Experiment 1 was designed to find the human brain areas that respond to a shift in location of the
221 form-defined FoE (i.e., encode form-defined FoEs). Specifically, we fixed the motion-defined FoE at the
222 center of the display (0°) and shifted the location of the form-defined FoE from -5° (left) to 5° (right) in
223 steps of 2°, resulting in six stimuli (Figure 2a). For each ROI, we thus trained a six-way linear SVM
224 classifier to discriminate the patterns of BOLD responses to the six stimuli. Figure S1 plots the decoding
225 accuracy as a function of the number of the most activated voxels (starting from 50 to the number that
226 covers the minimum number of the activated voxels across all participants) for each ROI. For all ROIs,
227 the decoding accuracy is stabilized at the voxel number of ≥ 100 . The white bars in Figure 2b thus plot the
228 classifier's decoding accuracy for each ROI with 100 voxels.



229

230

231

232

233

234

235

236

237

238

239

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 and positive sign indicates the FoE location to the right of the display center. The “x” and the “+” indicate the form- and the motion-defined FoEs, respectively. (b) The classifier’s decoding accuracy for the six stimuli for each ROI group. The white bars are for the voxel number of 100 and the gray bars are for the voxel number of 200. The dotted lines represent the 95th percentile of the classifier’s baseline decoding accuracies after 1000 shuffled tests. The solid line represents the chance level. The 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

240 accuracies than the baseline levels across 14 participants ($t(13) > 2.16, p < 0.05$). (d) The
241 classifier's decoding accuracy as a function of the difference in the form-defined FoEs.
242 The solid lines indicate the fitted linear functions. The error bars are SEs across 14
243 participants.

244 We grouped the ROIs as the early visual areas (V1, V2), the ventral visual areas (V3v, hV4, LO), the
245 dorsal visual areas (V3d, V3a, V7, V3B/KO), the dorsal motion visual areas (hMT, hMST), and other
246 optic flow areas (VIP, V6, CSv). We conducted a two-way (ROI \times decoding vs. baseline decoding
247 accuracy) repeated-measures ANOVA for each group and found that the classifier's decoding accuracy
248 was significantly higher than its baseline decoding accuracy for the early ($F(1,13) = 8.91, p = 0.011, \eta^2 =$
249 0.23), the ventral ($F(1,13) = 7.86, p = 0.015, \eta^2 = 0.1$), and the dorsal ($F(1,13) = 7.13, p = 0.019, \eta^2 =$
250 0.35) visual areas. No such main effect was found for the dorsal motion visual areas ($F(1,13) = 1.22, p =$
251 $0.29, \eta^2 = 0.086$) or other optic flow areas ($F(1,13) = 0.18, p = 0.68, \eta^2 = 0.014$). Tukey HSD tests
252 revealed that the classifier's decoding accuracy was significantly higher than its baseline accuracy for
253 areas V1 ($p = 0.0016$), V2 ($p = 0.00025$), V3v ($p = 0.00029$), V3d ($p = 0.02$), V3a ($p = 0.001$), and
254 V3B/KO ($p = 0.0011$), indicating that the pattern of BOLD responses of these visual areas can be
255 modulated by the shift in location of the form-defined FoE in the display. Because the minimum number
256 of the activated voxels across participants was larger than 200 for all these areas (see Figure S1), we thus
257 also computed the classifier's decoding accuracies by selecting 200 most activated voxels in these areas
258 as plotted by the gray bars in Figure 2b. Separate paired t -tests showed that the decoding accuracy with
259 the voxel number of 200 was not significantly different from that with the voxel number of 100 for all
260 these areas ($t(13) < 1.94, p > 0.07, \text{Cohen's } d < 0.52$). Due to the fact that the classifier's decoding
261 sensitivity in general increases with the number of selected voxels, in the following analyses, we trained
262 the classifier and computed its decoding accuracy by selecting 200 most activated voxels for these areas.

263 To examine whether any high-level brain areas also respond to the form-defined FoE shift, we
264 conducted searchlight MVPA analysis. The classifier's decoding accuracy was computed for the central
265 voxel of each spherical aperture, resulting in a map of decoding accuracy of the whole brain for each
266 participant. We set a cluster size threshold of 25 voxels and performed paired t -tests to compare each

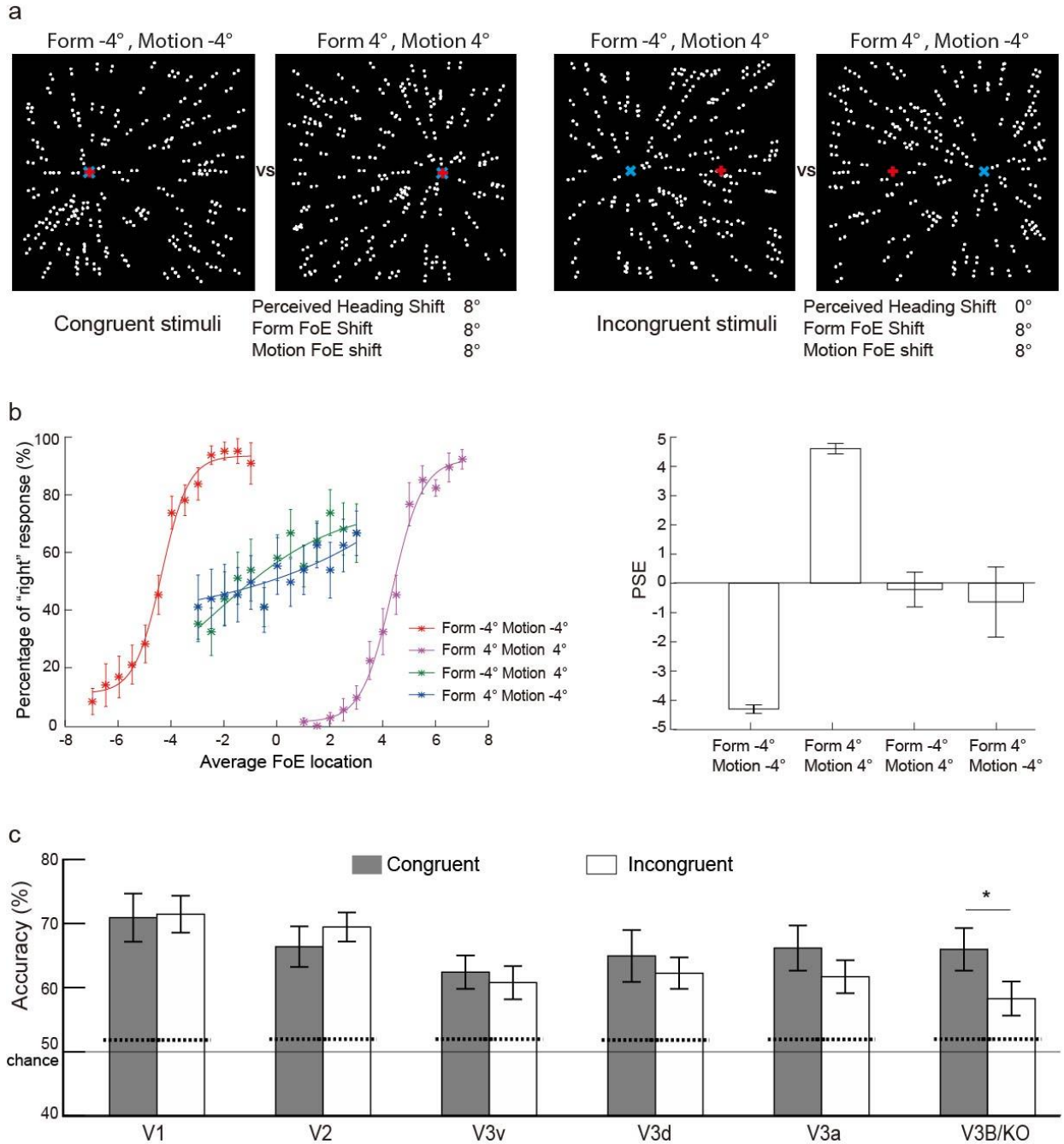
267 cluster's decoding accuracy with its baseline decoding accuracy across participants. We found that
268 consistent with the results of the ROI-based MVPA analysis, the early visual areas V1 and V2, the ventral
269 visual area V3v, and the dorsal visual areas V3d, V3a, and V3B/KO showed significantly higher decoding
270 accuracies than the baseline level. Furthermore, we did not observe any high-level brain areas involved in
271 decoding the form-defined FoE shift (Figure 2c).

272 How do these brain areas represent form-defined FoEs in heading perception? Do they only respond
273 to the form-defined FOE position shift or their response can be modulated by the magnitude of the
274 position shift? To address this question, we trained a two-way classifier with 200 voxels to discriminate
275 the pattern of BOLD responses when the difference in the form-defined FoEs in the six stimuli was 2°, 4°,
276 6°, 8°, or 10°. Figure 2d plots the classifier's decoding accuracy as a function of the difference in the
277 form-defined FoEs for these brain areas. A simple linear regression analysis revealed a significant linear
278 trend between the decoding accuracy and the difference in the form-defined FoEs for areas V3v ($R^2 =$
279 $0.93, p = 0.005$), V3d ($R^2 = 0.74, p = 0.04$), and V3a ($R^2 = 0.91, p = 0.008$) but not for areas V1 ($R^2 =$
280 $0.33, p = 0.18$), V2 ($R^2 = 0.43, p = 0.14$), and V3B/KO ($R^2 = -0.31, p = 0.84$). This suggests that while all
281 these six areas respond to the form-defined FOE position shift, only the responses in areas V3v, V3d, and
282 V3a can be modulated by the magnitude of the position shift.

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

289 *Areas integrating motion and form cues for heading perception*

290 In Experiment 2, we tested two types of stimuli in which the form- and the motion-defined FoE
291 locations were congruent (i.e., both were at -4° or 4°) or incongruent (i.e., the motion-defined FoE was at
292 -4° and the form-defined FoE was at 4° or vice versa, Figure 3a). Before scanning, we conducted the
293 psychophysical experiment to examine participants' heading perception. We found that for the two
294 congruent stimuli, the mean PSE averaged across 15 participants was $-4.34^\circ \pm 0.15^\circ$ (mean \pm SE) or
295 $4.61^\circ \pm 0.18^\circ$ when the motion- and the form-defined FoEs were at -4° or 4° . For the two incongruent
296 stimuli, the mean PSE was $-0.24^\circ \pm 0.59^\circ$ or $-0.66^\circ \pm 1.21^\circ$ when the motion-defined FoE was at 4° and the
297 form-defined FoE was at -4° or vice versa (Figure 3b). Separate paired *t*-tests revealed that while the
298 mean PSE was significantly different for the two congruent stimuli ($t(14) = -50.84$, $p < 0.001$, Cohen's $d =$
299 -13.13) but not for the two incongruent stimuli ($t(14) = 0.15$, $p = 0.89$, Cohen's $d = 0.038$). This indicates
300 that the perceived direction of heading shifted with the congruent but not with the incongruent stimuli.
301 Due to the fact that the change in motion and form signals in the two congruent stimuli was the same as in
302 the two incongruent stimuli, the brain areas that show a higher decoding accuracy for the congruent than
303 the incongruent stimuli should be responding to the perceived direction of heading rather than the change
304 in motion or form signals.



305
306
307
308
309
310
311
312
313
314

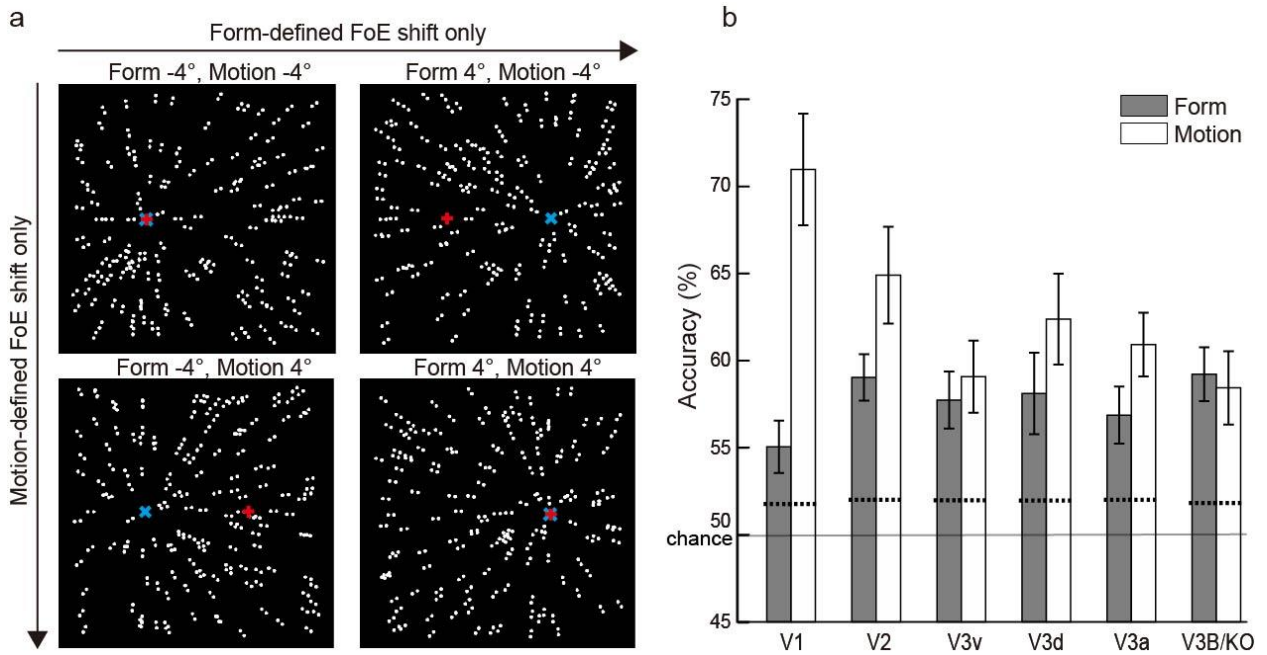
Figure 3. Experiment 2 visual stimuli and data. (a) Illustrations of the four stimuli. Negative sign indicates the FoE location to the left of the display center and positive sign indicates the FoE location to the right of the display center. The "x" and the "+" indicate the form- and the motion-defined FoEs, respectively. (b) Data from the psychophysical experiment. Left panel: 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 panel: Mean PSE against the four stimuli. Error bars are SEs across 15 participants. (c) The classifier's decoding accuracy for the congruent (gray)

315 and incongruent (white) stimuli for the six visual areas that respond to the form-defined
316 FoE shift in Experiment 1. The dotted lines represent the 95th percentile of the
317 classifier's baseline decoding accuracies. The solid line represents the chance level. The
318 error bars indicate SEs across 13 participants. *: $p < 0.05$.

319 Following this logic, we trained a two-way classifier to discriminate the pattern of BOLD responses
320 for the two congruent stimuli and the two incongruent stimuli, respectively. Figure 3c plots the classifier's
321 decoding accuracy along with the 95th percentile of the classifier's baseline decoding accuracy for the
322 congruent and incongruent stimuli for the six visual areas identified in Experiment 1. A 2 (decoding vs.
323 baseline decoding accuracy) x 2 (congruent vs. incongruent stimuli) repeated-measures ANOVA revealed
324 that for areas V1, V2, V3v, V3d, and V3a, only the main effect of decoding accuracy was significant
325 ($F(1,12) > 22.04$, $p < 0.00052$, $\eta^2 > 0.65$). For area V3B/KO, both the main effects of decoding accuracy
326 and stimulus type as well as their interaction effect were significant ($F(1,12) = 21.32$, $p = 0.0006$, $\eta^2 =$
327 0.64 , $F(1,12) = 6.63$, $p = 0.024$, $\eta^2 = 0.36$, and $F(1,12) = 6.72$, $p = 0.024$, $\eta^2 = 0.36$, respectively). Tukey
328 HSD tests showed that the decoding accuracy for the two congruent or incongruent stimuli was
329 significantly higher than its corresponding baseline level for all six visual areas ($p < 0.0092$), indicating
330 that these areas can discriminate the two stimuli of either the congruent or incongruent type. Nevertheless,
331 while there was no significant difference in the classifier's decoding accuracy between the congruent and
332 the incongruent stimulus types for areas V1 ($p = 0.997$), V2 ($p = 0.52$), V3v ($p = 0.86$), V3d ($p = 0.67$),
333 and V3a ($p = 0.38$), the classifier's decoding accuracy was significantly higher for the congruent than the
334 incongruent stimulus type for area V3B/KO ($p = 0.015$). This suggests that area V3B/KO plays an
335 important role in the integration of motion and form signals for the perception of heading.

336 Both the motion- and the form-defined FoEs changed their locations in the two congruent and the two
337 incongruent stimuli. The higher than baseline decoding accuracy observed for both types of stimuli thus
338 does not tell us whether the brain area responded to a shift in location of the motion- or the form-defined
339 FoE or both. To separate the brain area's responses to the motion- and the form-defined FoE shifts, we
340 examined the BOLD responses to the stimuli in which the shift in location only happened for the motion-
341 or the form-defined FoE (see Figure 4a). To illustrate, in the stimuli when only the location of the motion-

342 defined FoE was shifted, the form-defined FoE was fixed (at -4° or 4°) while the motion-defined FoE was
343 shifted from -4° to 4° . Similarly, in the stimuli when only the location of the form-defined FoE was
344 shifted, the motion-defined FoE was fixed (at -4° or 4°) while the form-defined FoE was shifted from -4°
345 to 4° . We trained a two-way classifier to discriminate the patterns of BOLD responses to the motion- or
346 the form-defined FoE shift. Figure 4b plots the classifier's decoding accuracy along with the
347 95th percentile of the classifier's baseline decoding accuracy for the form- and the motion-defined FoE
348 shifts for the six visual areas identified in Experiment 1. A 2 (decoding vs. baseline decoding accuracy) x
349 2 (form vs. motion cue) repeated-measures ANOVA revealed that while the main effect of decoding
350 accuracy was significant for all six visual area ($F(1,12) > 13.01, p < 0.0036, \eta^2 > 0.52$), the main effect of
351 cue type and the interaction effect of decoding accuracy and cue type were also significant for areas V1
352 ($F(1,12) = 12.73, p = 0.0039, \eta^2 = 0.52$ and $F(1,12) = 12.9, p = 0.0037, \eta^2 = 0.52$, respectively) and V2
353 ($F(1,12) = 5.9, p = 0.032, \eta^2 = 0.33$ and $F(1,12) = 6.23, p = 0.028, \eta^2 = 0.34$, respectively). Tukey HSD
354 tests revealed that the decoding accuracy for the motion- or the form-defined FoE shift was significantly
355 higher than the baseline level for all the visual areas except area V1 ($p < 0.038$), indicating that these
356 areas respond to either the motion- or the form-defined FoE shift. For area V1, the decoding accuracy for
357 the motion-defined FoE shift was significantly higher than the baseline level ($p = 0.00022$) and the
358 decoding accuracy for the form-defined FoE shift was only borderline significantly higher than the
359 baseline level ($p = 0.085$). This could be due to the fact that for both areas V1 and V2, the decoding
360 accuracy was significantly higher for the motion- than the form-defined FoE shift ($p < 0.019$), indicating
361 that these two areas have a higher response to the motion than the form information in the stimuli.



362

363 Figure 4. Visual stimuli and decoding accuracies for the motion- or the form-defined FoE
 364 shift. a) Illustrations of the stimuli with only the motion- or only the form-defined FoE
 365 shift. The “x” and the “+” indicate the form- and the motion-defined FoEs, respectively.
 366 b) The classifier’s decoding accuracy for the motion- (white) or the form-defined FoE
 367 shift (gray) for the six visual areas identified in Experiment 1. The dotted lines represent
 368 the 95th percentile of the classifier’s baseline decoding accuracies. The solid line
 369 represents the chance level. The error bars indicate SEs across 13 participants.

370 ***Neural computation for integrating motion and form cues***

371 How do the brain areas that encode either motion- or form-defined FoEs combine motion and form
 372 signals when they are presented simultaneously? There are two possibilities, linear optimal combination
 373 and fusion. For linear optimal combination, the brain area processes two types of cues as independent
 374 components and combines them in a statistically optimal manner according to the Bayes theorem (Landy
 375 et al., 1995; Ban et al., 2012). In this case, the classifier’s sensitivity to two consistent cues should be the
 376 quadratic sum of its sensitivity to each cue alone. In contrast, for fusion, the brain area may not process
 377 two types of cues independently and may also combine them in a nonlinear way (Ban et al., 2012). In this
 378 case, the classifier’s sensitivity to two consistent cues would not be equal to the quadratic sum of its
 379 sensitivity to each cue alone.

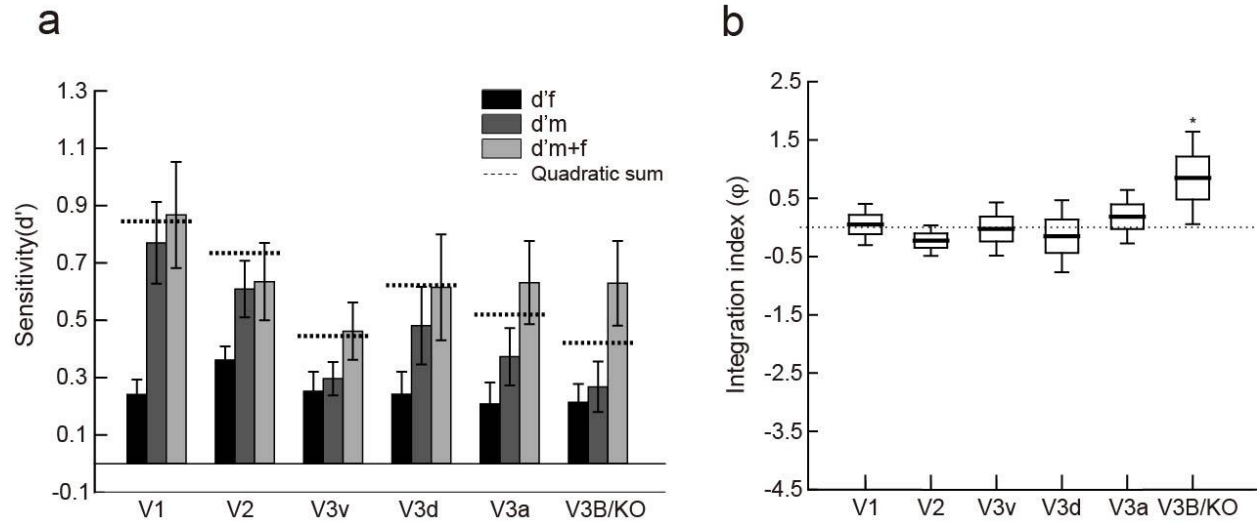
380 A classifier's sensitivity (d') to decode the neural responses to a cue can be computed using its
381 decoding accuracy for that cue (Ban et al., 2012):

$$382 \quad d' = 2\text{erf}^{-1}(2p - 1), \quad (1)$$

383 where p is the decoding accuracy. To examine how brain areas combine motion and form cues for
384 heading perception, we computed the classifier's form cue sensitivity index (d'_f) using its decoding
385 accuracy for only the form-defined FoE shift (gray bars, Figure 4b), the classifier's motion cue sensitivity
386 index (d'_m) using its decoding accuracy for only the motion-defined FoE shift (white bars, Figure 4b) and
387 the classifier's combined cue sensitivity index (d'_{m+f}) using its decoding accuracy for both the motion-
388 and the form-defined FoE shift in the two congruent stimuli (gray bars, Figure 3c). Figure 5a plots d'_f ,
389 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
390 make the comparison of d'_{m+f} to the quadratic sum of d'_m and d'_f easier, we converted the sensitivities
391 indices to an integration index (φ):

$$392 \quad \varphi = \frac{d'_{m+f}}{\sqrt{d_f'^2 + d_m'^2}} - 1. \quad (2)$$

393 Figure 5b plots the integration index for each visual area. Separate t -tests revealed that while the
394 integration index was significantly above zero for area V3B/KO ($t(12) = 2.31$, $p = 0.04$, Cohen's $d =$
395 0.64), it was not significantly different from zero for the other five areas (V1: $t(12) = 0.3$, $p = 0.77$,
396 Cohen's $d = 0.08$; V2: $t(12) = -1.86$, $p = 0.088$, Cohen's $d = -0.52$; V3d: $t(12) = -0.53$, $p = 0.61$, Cohen's
397 $d = -0.15$; V3a: $t(12) = 0.86$, $p = 0.41$, Cohen's $d = 0.24$). This suggests that in contrast to areas V1, V2,
398 V3v, V3d, and V3a that perform linear optimal combination when responding to motion and form cues,
399 area V3B/KO performs fusion computation when combining these two cues for heading perception.



400
401
402
403
404
405
406
407
408

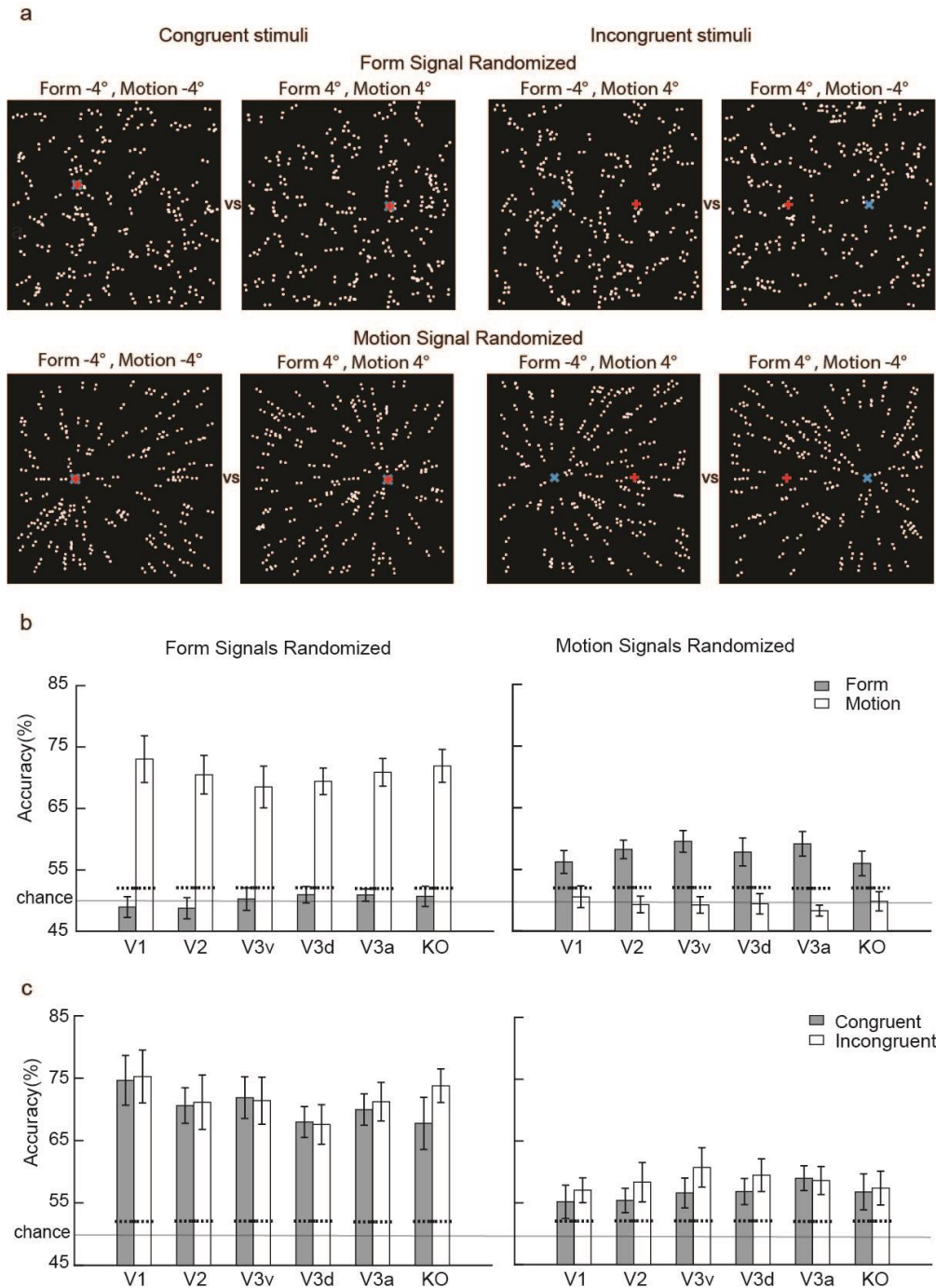
Figure 5. Sensitivity and integration index data. a) The motion cue (d'_m), the form cue (d'_f), and the combined cue (d'_{m+f}) sensitivity indices for the six visual areas that encode either the motion- or the form-defined FoE shift. The dotted lines represent the quadratic sums of d'_m and d'_f . The error bars indicate SEs across 13 participants. b) The integration index for the six visual areas. The black line in the center of each bar indicates the median, the edges depict 68% confidence intervals, and the error bars depict 95% confidence intervals. The dashed line at zero indicates the quadratic sum of d'_m and d'_f . *: $p < 0.05$.

409
410

Randomizing form or motion signals

411
412
413
414
415
416
417

To validate whether the responses in the cortical 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 display 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 (Figure 6a, top row), and randomizing the motion signals removed the motion-defined FoE but left the form-defined FoE intact (Figure 6a, bottom row).



418

419

420

421

422

423

424

425

426

Figure 6. Experiment 3 visual stimuli and data. (a) Illustrations of the form signal randomized motion signal intact stimuli (top row) and the motion signal randomized form signal intact stimuli (lower row). Negative sign indicates the FoE location to the left of the display center and positive sign indicates the FoE location to the right of the display center. The “x” and the “+” 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 form (left) and motion (right) signal randomized stimuli for the six visual areas identified in Experiment 1. (c) The classifier’s decoding accuracy

427 for the congruent (gray) and incongruent (white) stimuli the form (left) and motion
428 (right) signal randomized stimuli for the six visual areas. The dotted lines represent the
429 95th percentile of the classifier's baseline decoding accuracies. The solid line represents
430 the chance level. The error bars indicate SEs across 12 participants.

431 As in Experiment 2, we trained a two-way classifier to discriminate the patterns of BOLD responses
432 to the motion- or the form-defined FoE shift. Figure 6b plots the classifier's decoding accuracy along with
433 the 95th percentile of the classifier's baseline decoding accuracy for the form- and the motion-defined FoE
434 shifts for the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right). A 2
435 (decoding vs. baseline decoding accuracy) x 2 (form vs. motion cue) repeated-measures ANOVA
436 revealed that for both the form- and the motion-signal-randomized stimuli, the interaction effect of
437 decoding accuracy and cue type was significant for all six visual areas ($F(1,11) > 6.55, p < 0.027, \eta^2 >$
438 0.37). Tukey HSD tests showed that for the form-signal-randomized stimuli, while the decoding accuracy
439 for the motion-defined FoE shift was significantly higher than the baseline level for all six visual areas (p
440 < 0.00038), the decoding accuracy for the form-defined FoE shift was not different from the baseline
441 level for all six visual areas ($p > 0.91$). In contrast, for the motion-signal-randomized stimuli, while the
442 decoding accuracy for the form-defined FoE shift was significantly higher than the baseline level for all
443 six visual areas ($p < 0.01$), the decoding accuracy for the motion-defined FoE shift was not different from
444 the baseline level for all six visual areas ($p > 0.68$). This shows that randomizing the form signals to
445 remove the form cue to the FoE indeed affected the decoding accuracy for the form-defined FoE shift
446 only and randomizing the motion signals to remove the motion cue to the FoE indeed affected the
447 decoding accuracy for the motion-defined FoE shift only, thus supporting the claim that the responses in
448 the cortical areas identified in Experiments 1 and 2 are driven by global form and motion signals.

449 Because randomizing the form or the motion signals in the four stimuli of Experiment 2 removed the
450 change in the form or the motion signals thus making the two congruent stimuli the same as the two
451 incongruent stimuli, we expected that all the visual areas would show similar decoding accuracies for the
452 congruent and the incongruent stimuli. To examine this, as in Experiment 2, we trained a two-way
453 classifier to discriminate the pattern of BOLD responses for the two congruent stimuli and the two

454 incongruent stimuli, respectively. Figure 6c plots the classifier's decoding accuracy along with the
455 95th percentile of the classifier's baseline decoding accuracy for the congruent and incongruent stimuli for
456 the form-signal-randomized stimuli (left) and the motion-signal-randomized stimuli (right). Separate 2
457 (decoding vs. baseline decoding accuracy) x 2 (congruent vs. incongruent stimuli) repeated-measures
458 ANOVAs showed that for both the form- and the motion-signal-randomized stimuli, only the main effect
459 of decoding accuracy was significant for all six visual areas ($F(1,11) > 8.96$, $p < 0.012$, $\eta^2 > 0.45$), i.e.,
460 across the congruent and incongruent stimuli types, the decoding accuracy was significantly higher than
461 the baseline decoding accuracy for all six visual areas. Tukey HSD tests showed that for both the form-
462 and the motion-signal-randomized stimuli, there was no significant difference in the classifier's decoding
463 accuracy between the congruent and the incongruent stimuli for all six visual areas ($p > 0.12$). This
464 confirms that when randomizing the form or the motion signals to render the two congruent stimuli the
465 same as the two incongruent stimuli, all the visual areas identified in Experiments 1 and 2 indeed could
466 not tell the difference between the congruent and the incongruent stimuli any more.

467 *Eye movement data*

468 In all three brain-imaging experiments, on each trial, we presented a red fixation point at the center of
469 the display for 400 ms followed by the self-motion display for 600 ms. We did not present any fixation
470 point in the self-motion display to ensure that the self-motion display did not contain any extraneous
471 relative motion. We nevertheless instructed participants to maintain their eye position at the center of the
472 display throughout the trial. If participants followed our instructions, then the pattern of their eye
473 movements should not vary across the stimulus conditions in all experiments. To examine whether
474 participants could follow our instructions, in a separated session outside of the scanner, we recorded eye
475 movements of six participants who all participated in Experiments 1 and 2 when they viewed the same
476 display and performed the same task as in Experiments 1 and 2.

477 The recorded eye movement data are plotted in Figure S2. For Experiment 1, a one-way repeated-
478 measures ANOVA (with the Greenhouse–Geisser correction for any lack of sphericity) revealed no

479 significant difference in the horizontal ($F(5, 25) = 1.54, p = 0.21, \eta^2 = 0.24$) or vertical ($F(2.5, 12.5) =$
480 $0.28, p = 0.81, \eta^2 = 0.053$) eye positions across the six stimulus conditions. There was also no
481 significance difference in saccade amplitude ($F(5,25) = 0.39, p = 0.85, \eta^2 = 0.072$) or the number of
482 saccades ($F(1.55, 7.75) = 0.997, p = 0.39, \eta^2 = 0.17$) across the six stimuli. For Experiment 2, similarly, a
483 one-way repeated-measures ANOVA (with the Greenhouse–Geisser correction for any lack of sphericity)
484 revealed no significant difference in the horizontal ($F(1.01, 5.07) = 2.69, p = 0.16, \eta^2 = 0.35$) or vertical
485 ($F(3, 15) = 0.85, p = 0.49, \eta^2 = 0.15$) eye positions across the four stimulus conditions. There was also no
486 significance difference in saccade amplitude ($F(3, 15) = 1.95, p = 0.17, \eta^2 = 0.28$) or the number of
487 saccades ($F(3, 15) = 1.59, p = 0.23, \eta^2 = 0.24$) across the four stimuli. These results support the claim that
488 participants were able to follow the instructions and maintain their eye at the center of the display
489 throughout the trial.

490 **Discussion (1441/1500)**

491 Combining the results from the three experiments, we found that the early visual areas V1, V2, and
492 V3 (V3v and V3d combined) respond to a position shift of the FoE defined by either motion or form cues.
493 This is consistent with the findings of primate neurophysiology studies showing that these areas process
494 both local motion and form information (Hubel and Wiesel, 1968; Mikami et al., 1986; Felleman and Van
495 Essen, 1987; Levitt et al., 1994; Gegenfurtner et al., 1997; Hu et al., 2018). Research identifying the
496 homology of primate areas V1, V2, and V3 in the human brain has been quite successful and shows that
497 these areas in humans are organizationally and functionally analogous to those in macaques. However, for
498 visual areas beyond V3, the homology between the primate and human brain breaks down and is less
499 certain (Winawer and Witthoft, 2015).

500 Our results show that after area V3, the dorsal (V3a and V3B/KO) rather than the ventral visual areas
501 (hV4 and LO) respond to either the motion- or the form-defined FOE shift. Previous research has shown
502 that the center of a radial flow pattern activates area V3a, suggesting that this area responds to the exact
503 location of the FoE in optic flow (Koyama et al., 2005). Our finding regarding area V3a thus

504 complements previous findings for this area. Our findings are also consistent with the dissociation of the
505 ventral and dorsal streams regarding visual information processing for perception and action (Goodale
506 and Milner, 1992). Specifically, the ventral stream recognizes and discriminates shape, size, and color of
507 objects (Kravitz et al., 2013) and thus supports vision for perception, whereas the dorsal stream encodes
508 spatial location, orientation, and motion of objects to guide actions and thus supports vision for action
509 (Decety and Grezes, 1999). Because our stimuli provide heading information that can be used for the
510 control of self-motion (e.g., Gibson, 1950; L. Li and Niehorster, 2014), it is reasonable that the dorsal but
511 not ventral visual areas respond to the motion- or the form-defined FoE shift.

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

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

539 Previous studies have shown that area V3B/KO is a candidate cortical locus for the integration of
540 qualitatively different cues. For example, Ban et al. (2012) found that area V3B/KO integrates disparity
541 and motion information for depth perception. It has been further revealed that the cue integration in area
542 V3B/KO is not specific to specific cue pairing (such as disparity and motion) but can be generalized to
543 different cue pairings, such as disparity and shading (Dovencioğlu et al., 2013) or disparity and texture
544 (Murphy et al., 2013). Using transcranial direct current stimulation to perturb the excitatory and inhibitory
545 balance of area V3B/KO leads to impaired performance of such cue integration (Rideaux and Welchman,
546 2018). Our study used quite different types of stimuli from those previously used, the motion- and form-
547 defined FoEs, that are also qualitatively different. The integration of these cues in area V3B/KO for the
548 perception of heading is thus compatible with previous findings and suggests quite general integration
549 computations with area V3B/KO that could not be inferred from previous studies.

550 The results of the current study show that neither the dorsal motion (MT and MST) nor other optic
551 flow visual areas (VIP, V6, and CSv) are involved in the integration of motion and form cues for the
552 perception of heading. While we do not exclude the possibility that this could be due to the sampling and

553 measurement approach we took in the current study², we believe that this is more related to the ability to
554 encode fine differences in the FoE location using the activity of spatially-precise receptive fields in early
555 visual areas. For example, studies have shown that the human homologue of primate MST can
556 discriminate expansion from contraction flow patterns but does not appear to encode the specific location
557 of the FoE in optic flow (Strong et al., 2017), and human V6 is also not sensitive to the change in location
558 of the FoE in optic flow (Furlan et al., 2013). In addition, previous findings of primate neurophysiology
559 studies show that most MST neurons do not respond to form information (Geesaman and Andersen,
560 1996), and area VIP receives a large amount of input from area MST but not much input from the ventral
561 stream (Ungerleider et al., 2008). In contrast to other flow selective brain areas, CSv responses to optic
562 flow can be suppressed by many factors such as whether the flow pattern is compatible with self-motion
563 (Wall and Smith, 2008) or whether flow is used for visuomotor control (Field et al., 2015). All these
564 factors can contribute to the lack of responses in higher visual areas associated with optic flow processing
565 to the form-defined FoE shift and thus the lack of involvement in the integration of motion and
566 form cues for the perception of heading.

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

² 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 does not respond to the form-defined FoE shift and thus is not involved in the integration of motion and form cues for the perception of heading.

572 /

573 **References**

574 Ban H, Preston TJ, Meeson A, Welchman AE (2012) The integration of motion and disparity cues to
575 depth in dorsal visual cortex. *Nat Neurosci* 15(4):636–643.

576 Beer J, Blakemore C, Previc FH, Liotti M (2002) Areas of the human brain activated by ambient visual
577 motion, indicating three kinds of self-movement. *Exp Brain Res* 143(1):78–88.

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

580 Braddick O, O'Brien J, Wattam-Bell J, Atkinson J, Turner R (2000) Form and motion coherence activate
581 independent, but not dorsal/ventral segregated, networks in the human brain. *Curr Biol* 10(12):731–
582 734.

583 Burr DC, Ross J (2002) Direct evidence that “speedlines” influence motion mechanisms. *J Neurosci* 22:
584 8661–8664.

585 Crowell JA, Banks MS (1993) Perceiving heading with different retinal regions and types of optic
586 flow. *Percept Psychophys* 53(3):325–337.

587 Decety J, Grezes J (1999) Neural mechanisms subserving the perception of human actions. *Trends Cogn*
588 *Sci* 3(5):172–178.

589 DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. *Trends*
590 *Neurosci* 11(5):219–226.

591 DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser JON, Cox R, Miller D, Neitz J (1996)
592 Mapping striate and extrastriate visual areas in human Cereb Cortex. *Proc Natl Acad Sci*
593 *USA* 93(6):2382–2386.

594 Dovencioğlu D, Ban H, Schofield AJ, Welchman AE (2013) Perceptual Integration for Qualitatively
595 Different 3-D Cues in the Human Brain. *J Cogn Neurosci* 25(9):1527–1541.

596 Dukelow SP, DeSouza JFX, Culham JC, van den Berg AV, Menon RS, Vilis T (2001) Distinguishing
597 subregions of the human MT plus complex using visual fields and pursuit eye movements. *J*
598 *Neurophysiol* 86(4):1991–2000.

599 Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marchal G, Mortelmans L, Orban GA
600 (1997) The kinetic occipital region in human visual cortex. *Cereb Cortex* 7(3):283–292.

601 Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI
602 of human visual cortex. *Nature* 369(6481):525.

603 Felleman DJ, Van Essen DC (1987) Receptive field properties of neurons in area V3 of macaque monkey
604 extrastriate cortex. *J Neurophysiol* 57(4):889–920.

605 Field D, Inman LA, Li L (2015) Visual processing of optic flow and motor control in the human posterior
606 cingulate sulcus. *Cortex* 71, 377-389.

607 Furlan M, Wann JP, Smith AT (2013) A representation of changing heading direction in human cortical
608 areas pVIP and CSv. *Cereb Cortex* 24(11):2848–2858.

609 Geesaman BJ, Andersen RA (1996) The analysis of complex motion patterns by form/cue invariant
610 MSTd neurons. *J Neurosci* 16(15):4716–4732.

611 Gegenfurtner KR, Kiper DC, Levitt JB (1997) Functional properties of neurons in macaque area V3. *J*
612 *Neurophysiol* 77(4):1906–1923.

613 Geisler WS (1999) Motion streaks provide a spatial code for motion direction. *Nature* 400(6739):65–69.

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

615 Glass L (1969) Moire effect from random dots. *Nature* 223(5206):578–580.

616 Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci*
617 15(1):20–25.

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

620 Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque
621 MSTd. *Nat Neurosci* 11(10):1201–1210.

622 Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary
623 visual cortex. *Nat Neurosci* 8(5):686–691.

624 Hu JM, Ma H, Zhu SD, Li PC, Xu HR, Fang Y, Chen M, Han C, Fang C, Cai XY, et al (2018) Visual
625 motion processing in macaque V2. *Cell Rep* 25:157–167.

626 Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J*
627 *Physiol* 195(1):215–243.

628 Johansson G (1973) Visual perception of biological motion and a model for its analysis. *Percept*
629 *Psychophys* 14(2):201–211.

630 Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci*
631 8(5):679–685.

632 Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital
633 complex. *Science* 293(5534):1506–1509.

634 Kourtzi Z, Krekelberg B, van Wezel RJA (2008) Linking form and motion in the primate brain. *Trends*
635 *Cogn Sci* 12(6):230–236.

636 Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: an
637 expanded neural framework for the processing of object quality. *Trends Cogn Sci* 17(1):26–49.

638 Krekelberg B, Vatakis A, Kourtzi Z (2005) Implied motion from form in the human visual cortex. *J*
639 *Neurophysiol* 94(6):4373–4386.

640 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl*
641 *Acad Sci USA* 103(10):3863–3868.

642 Koyama S, Sasaki Y, Andersen GJ, Tootell RB, Matsuura M, Watanabe T (2005) Separate processing of
643 different global-motion structures in visual cortex is revealed by fMRI. *Curr Biol* 15(22):2027–2032.

644 Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue
645 combination: in defense of weak fusion. *Vision Res* 35(3):389–412.

646 Levitt JB, Kiper DC, Movshon JA (1994) Receptive fields and functional architecture of macaque V2. *J*
647 *Neurophysiol* 71(6):2517–2542.

648 Li L, Niehorster DC (2014) Influence of optic flow on the control of heading and target egocentric
649 direction during steering toward a goal. *J Neurophysiol* 112(4), 766-777.

650 Li L, Peli E, Warren WH (2002) Heading perception in patients with advanced retinitis pigmentosa.
651 *Optom Vis Sci* 79(9):581–589.

652 Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain.
653 *J Neurosci* 27(45):12321–12330.

654 Mikami A, Newsome WT, Wurtz RH (1986) Motion selectivity in macaque visual cortex. II.
655 Spatiotemporal range of directional interactions in MT and V1. *J Neurophysiol* 55(6):1328–1339.

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

658 Murphy AP, Ban H, Welchman AE (2013) Integration of texture and disparity cues to surface slant in
659 dorsal visual cortex. *J Neurophysiol* 110(1):190–203.

660 Niehorster, DC, Cheng, JC, Li L (2010) Optimal combination of form and motion cues in human heading
661 perception. *J Vis* 10(11):20,1–15.

662 Orban GA, Claeys K, Nelissen K, Smans R, Sunaert S, Todd JT, Wardak C, Durand JB, Vanduffel W
663 (2006) Mapping the parietal cortex of human and non-human primates. *Neuropsychologia*
664 44(13):2647–2667.

665 Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys
666 and humans. *Trends Cogn Sci* 8(7):315–324.

667 Ostwald D, Lam JM, Li S, Kourtzi Z (2008) Neural coding of global form in the human visual cortex. *J*
668 *Neurophysiol* 99(5):2456–2469.

669 Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Patria F, Galletti C (2010) Human V6: the
670 medial motion area. *Cereb Cortex* 20(2):411–424.

671 Rideaux R, Welchman AE (2018) Proscription supports robust perceptual integration by suppression in
672 human visual cortex. *Nat Commun* 9(1):1502.

673 Rutschmann RM, Schrauf M, Greenlee MW (2000) Brain activation during dichoptic presentation of
674 optic flow stimuli. *Exp Brain Res* 134(4):533–537.

675 Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995)
676 Borders of multiple visual areas in humans revealed by functional magnetic resonance
677 imaging. *Science* 268(5212):889–893.

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

680 Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, Pestilli F (2016) A major human white
681 matter pathway between dorsal and ventral visual cortex. *Cereb Cortex* 26(5):2205–2214.

682 Ungerleider LG, Galkin TW, Desimone R, Gattass R (2008) Cortical connections of area V4 in the
683 macaque. *Cereb Cortex* 18(3):477–499.

684 van den Berg AV (1992) Robustness of perception of heading from optic flow. *Vision Res* 32(7):1285–
685 1296.

686 Wall MB, Smith AT (2008) The representation of egomotion in the human brain. *Curr Biol* 18(3):191–
687 194.

688 Wallach H, O’Connell DN (1953) The kinetic depth effect. *J Exp Psychol* 45(4):205–217.

689 Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56(2):366–
690 383.

691 Warren WH, Morris MW, Kalish M (1988) Perception of translational heading from optical flow. *J Exp*
692 *Psychol Hum Percept Perform* 14(4):646–660.

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

694 Yeatman JD, Weiner KS, Pestilli F, Rokem A, Mezer A, Wandell BA (2014) The vertical occipital
695 fasciculus: a century of controversy resolved by in vivo measurements. *Proc Natl Acad Sci*
696 *USA* 111(48): E5214–E5223.

697 Zeki S (2003) Improbable areas in the visual brain. *Trends Neurosci* 26(1):23–26.

698 Zeki S, Perry RJ, Bartels A (2003) The processing of kinetic contours in the brain. *Cereb Cortex*
699 13(2):189–202.

700 Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of
701 functional specialization in human visual cortex. *J Neurosci* 11(3):641–649.

702 Zihl J, Von Cramon D, Mai N (1983) Selective disturbance of movement vision after bilateral brain
703 damage. *Brain* 106(2):313–340.