

3D shape perception from combined depth cues in human visual cortex

Andrew E Welchman¹, Arne Deubelius¹, Verena Conrad¹, Heinrich H Bülhoff¹ & Zoe Kourtzi^{1,2}

Our perception of the world's three-dimensional (3D) structure is critical for object recognition, navigation and planning actions. To accomplish this, the brain combines different types of visual information about depth structure, but at present, the neural architecture mediating this combination remains largely unknown. Here, we report neuroimaging correlates of human 3D shape perception from the combination of two depth cues. We measured fMRI responses while observers judged the 3D structure of two sequentially presented images of slanted planes defined by binocular disparity and perspective. We compared the behavioral and fMRI responses evoked by changes in one or both of the depth cues. fMRI responses in extrastriate areas (hMT+/V5 and lateral occipital complex), rather than responses in early retinotopic areas, reflected differences in perceived 3D shape, suggesting 'combined-cue' representations in higher visual areas. These findings provide insight into the neural circuits engaged when the human brain combines different information sources for unified 3D visual perception.

Visual environments are defined by multiple cues to depth structure, such as binocular disparity, perspective, texture, shading and motion. The computations and neural mechanisms required to extract this information from the retinal images are likely to be radically different, but despite this, we perceive coherent 3D structures; somehow, the information provided by different depth cues is combined by the brain¹. The stereogram in **Figure 1a** provides an illustration of the combination process: in this 3D shape defined by two planes slanted in depth, horizontal binocular disparity and perspective cues provide different information as to the object's 3D structure. However, when observers view such stimuli, they combine the slant information from each cue to perceive an intermediate 3D shape. This is not merely a laboratory curiosity produced by inducing cue conflict; rather, the brain routinely combines different depth cues to obtain different types of information and reduce the effects of sensor and processing noise¹⁻³.

Recent studies have examined the neural mechanisms that mediate processing of individual depth cues: namely, disparity⁴⁻⁶ and perspective^{7,8}. This work has considered relatively isolated depth cues or multiple correlated cues, providing insights into sites where information about individual depth cues may converge. However, the neural substrates underlying the perception of shape based on the combination of cues to depth structure have not been investigated. We addressed this question using concurrent psychophysical and fMRI measurements.

We used a 3D shape stimulus depicting a hinged plane receding in depth (**Fig. 1**) in which the angular slant specified by two cues, horizontal binocular disparity and linear perspective, was different ('inconsistent-cue stimulus'). Observers judged the 3D shape of this test stimulus through comparisons with reference stimuli in which the

cues defining 3D structure were consistent ('consistent-cue stimuli'). In each trial, the test stimulus and a reference stimulus were presented sequentially, and they differed in one or both of the depth cues.

The fMRI measurements employed an event-related adaptation procedure^{9,10}. The technique capitalizes on neural adaptation and repetition suppression effects whereby neural activity is lower for stimuli that have been viewed recently than for stimuli that have not. By comparing the adapted response evoked by the test stimulus shown twice ('test-test') with responses evoked by the test stimulus paired with a reference stimulus ('test-reference'), we could deduce sensitivity in the neural population to differences between the test and reference. For example, a higher fMRI response (rebound effect) when the test-reference pair differs only in disparity information from the adapted test-test response would indicate sensitivity to changes in the disparity cue.

By comparing psychophysical behavior and fMRI rebound effects, we demonstrate that responses in extrastriate ventral and dorsal areas, rather than in early retinotopic areas, correspond to changes in perceived 3D shape based on cue combination. These findings provide evidence that higher visual areas are involved in processing perceived global shape that depends on the combination of individual depth cues.

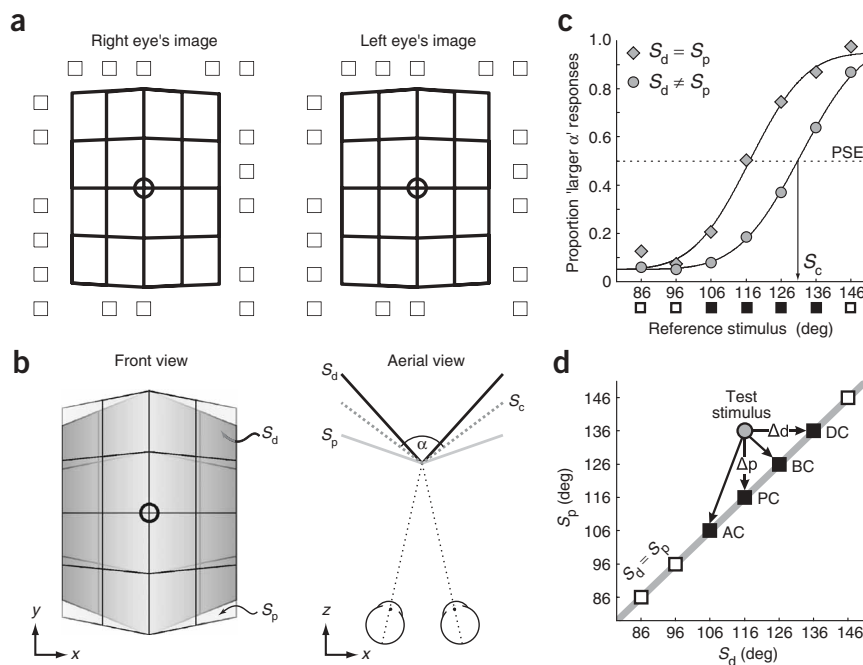
RESULTS

Experiment 1: cue-based and percept-based processing

Observers judged which of the two viewed stimuli in a trial had a larger dihedral angle (α : openness; **Fig. 1b**). By presenting the inconsistent-cue test stimulus (perspective angle, $S_p = 136^\circ$; disparity angle, $S_d = 116^\circ$) and a range of consistent-cue reference stimuli ($S_d = S_p$) we

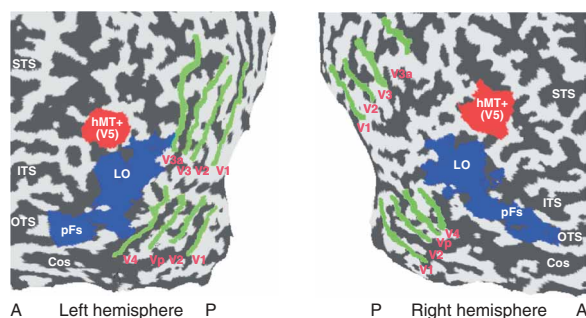
¹Max-Planck Institute for Biological Cybernetics, Postfach 2169, 72012 Tübingen, Germany. ²School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. Correspondence should be addressed to Z.K. (zoe.kourtzi@tuebingen.mpg.de).

Figure 1 Stimulus and design. (a) Stereogram of a stimulus similar to those used. Cross-eyed fusion of the images yields the impression of a hinged plane receding in depth. The perspective cue is provided by the trapezoidal projection of the rectangular plane; the disparity cue by the different positions of corresponding features in the two eyes. Moving the figure closer and further away manipulates the conflict between the cues. When the figure is moved away, the perceived angle between the planes gets smaller: the disparity cue specifies a more acute angle as distance increases, but the perspective-specified angle remains constant. (b) Cartoon portraying disparity-defined shape (S_d) as more acute than perspective-defined shape (S_p). Aerial view at right: the observer fixates the vertex, obtaining perspective (S_p) and disparity (S_d) information. Combining the cues results in perceived 3D shape (S_c) between S_d and S_p . (c) Psychometric functions ($n = 11$) for consistent-cue ($S_d = S_p = 116^\circ$) and inconsistent-cue ($S_d = 116^\circ$, $S_p = 136^\circ$) test stimuli. (d) Stimulus space. The diagonal $S_d = S_p$ refers to consistent-cue stimuli. The inconsistent-cue test stimulus (filled circle, $S_d = 116^\circ$, $S_p = 136^\circ$) was paired with consistent-cue reference stimuli (filled squares: $S_d = S_p = 136^\circ$, 126° , 116° , and 106°). DC, Disparity Change (test and reference differ in disparity: $\Delta S_d = 20^\circ$); BC, Balanced Change (perspective and disparity cues differ in opposite directions: $\Delta S_d = 10^\circ$, $\Delta S_p = -10^\circ$); PC, Perspective Change (test and reference differ in perspective: $\Delta S_p = -20^\circ$); AC, All Change (both cues differ: $\Delta S_d = -10^\circ$, $\Delta S_p = -30^\circ$). Open squares: consistent-cue reference stimuli used in psychophysical tests (but not in fMRI experiments).



obtained a psychometric function (Fig. 1c). Fitting these data yielded the point of subjective equality (PSE; 50% point on the curve), to provide a measure of perceived angle in the inconsistent-cue test stimulus (S_c). Observers combined the cues, perceiving an angle between that specified by each cue. However, the PSE ($130.8 \pm 0.4^\circ$) was closer to S_p (136°) than S_d (116°), indicating that perception was determined more by perspective than by disparity¹¹ (cue weights: $w_p = 0.73$; $w_d = 0.27$). As expected¹, there was some between-subjects variability in cue weights. However, all observers showed evidence of cue combination (Supplementary Table 1).

The fMRI analysis considered responses in independently defined regions of interest (ROIs; Fig. 2) within the visual cortex. We studied early retinotopic areas (V1, V2, V3, V3a, Vp, V4) involved in processing visual features. We also investigated two higher visual areas implicated in analyzing 3D information: the lateral occipital complex (LOC), which is involved in processing object shape¹², and hMT+/V5, which is implicated in the analysis of motion^{13,14}, structure from motion and depth^{15,16}.



To assess sensitivity to both individual cue changes and changes in perceived 3D shape, we used five experimental conditions (Fig. 1d): (i) the Identical Condition, in which the inconsistent-cue test stimulus was shown twice; (ii) the Disparity Change condition, in which the test-reference pair specified the same perspective information but different disparities; (iii) the Perspective Change condition, in which the perspective cue differed between test and reference; (iv) the Balanced Change condition, in which the cues changed in opposite directions and (v) the All Change condition, in which the cues changed in the same direction. For each individual subject, we computed the fMRI time course from each condition and ROI and normalized it to the response from the fixation baseline (Fig. 3a). We compared averaged fMRI responses at the peak of the fMRI time course in each ROI across conditions and subjects (Fig. 3b). The Identical Condition evoked significantly lower fMRI responses than the other conditions across visual areas ($F_{28,280} = 3.16$, $P < 0.05$), providing evidence for fMRI adaptation^{9,10}. Furthermore, we observed a significant interaction ($F_{4,280} = 2.29$, $P < 0.05$) for condition and ROI.

Figure 2 Regions of interest within the visual cortex. Functional activation maps for one subject showing the early retinotopic regions, hMT+/V5 and the LOC (comprising the regions labeled pFs and LO). The functional regions are superimposed on flattened cortical surfaces of the right and left hemispheres. Dark gray, sulci; light gray, gyri. A, anterior; P, posterior. STS: superior temporal sulcus, ITS: inferior temporal sulcus, OTS: occipitotemporal sulcus, CoS: collateral sulcus. The LOC was defined as the voxels in ventral occipitotemporal cortex showing significantly stronger activation ($P < 10^{-4}$, corrected) to intact images than to scrambled images of objects. hMT+/V5 was defined as the contiguous voxels in the ascending limb of the inferior temporal sulcus showing significantly stronger activation ($P < 10^{-4}$, corrected) to moving than to static low-contrast concentric rings. The borders of early retinotopic regions (V1, V2, V3, V3a, Vp, V4) were localized using rotating wedge stimuli, and eccentricity mapping was achieved using concentric rings.

To determine whether cortical areas process cue-based information, we compared **Table 1a** fMRI responses in the Identical Condition with those evoked when the presented stimuli differed in one cue (Disparity Change and Perspective Change conditions). The Disparity Change condition evoked significantly higher fMRI responses than the Identical Condition in V1, V2, V3, V3a, Vp and V4. The Perspective Change condition evoked significantly higher fMRI responses than the Identical Condition in V1, V2 and dorsal (V3, V3a) extrastriate areas; higher mean responses in ventral areas (Vp, V4) were not statistically significant.

In higher areas (LOC and hMT+/V5), the Perspective Change, but not the Disparity Change, condition evoked significantly higher activity than the Identical Condition. The small numerical trend for higher responses in the Disparity Change compared with the Identical condition (**Fig. 3b**) appears consistent with previous neurophysiological¹⁷ and fMRI adaptation¹⁸ studies showing sensitivity to disparity in higher visual areas. An interesting possibility is that the lack of strong sensitivity to disparity changes in our study is consistent with the psychophysical data, where perception was determined more by perspective than by disparity information.

To examine the relationship between perception and fMRI responses, we made comparisons based on psychophysical performance (**Fig. 3c**, inset). First, we compared fMRI responses in 'high-discriminability' (Perspective Change and All Change) and 'low-discriminability' (Balanced Change and Disparity Change)

conditions. We reasoned that areas whose activity reflects perceived 3D shape based on cue combination should show stronger fMRI responses in high-discriminability conditions than in low-discriminability conditions. This activation pattern was evident in areas LOC and hMT+/V5 but not in the retinotopic visual areas, with the exception of area V3a (**Fig. 3c**, **Table 1a**).

Second, we compared fMRI responses in conditions that had similar perceptual discriminability. We reasoned that when differences in perceived shape were similar, there should be no differences in fMRI response in areas that represent 3D shape based on the combination of cues. Based on psychophysics, the Balanced Change and Disparity Change conditions were not differentially discriminable (repeated measures ANOVA: $F_{1,30} < 0.01$, $P = 0.99$). Accordingly, fMRI responses in the Balanced Change and Disparity Change conditions were not significantly different in the higher visual areas (LOC, hMT+/V5) and in area V3a. In contrast, the Disparity Change condition evoked significantly higher fMRI responses than the Balanced Change condition in the other retinotopic visual areas (V1, V2, V3, Vp, V4). Taken together, these dissociable results suggest sensitivity to cue changes in the retinotopic visual areas, excepting V3a, whereas responses in extrastriate ventral (LOC) and dorsal (hMT+/V5) visual areas reflect differences in perceived 3D shape that are dependent on cue combination.

To quantify the relationship between psychophysical and fMRI responses, we conducted regression analyses (**Fig. 4**), which provided evidence for a significant relationship between psychophysical and

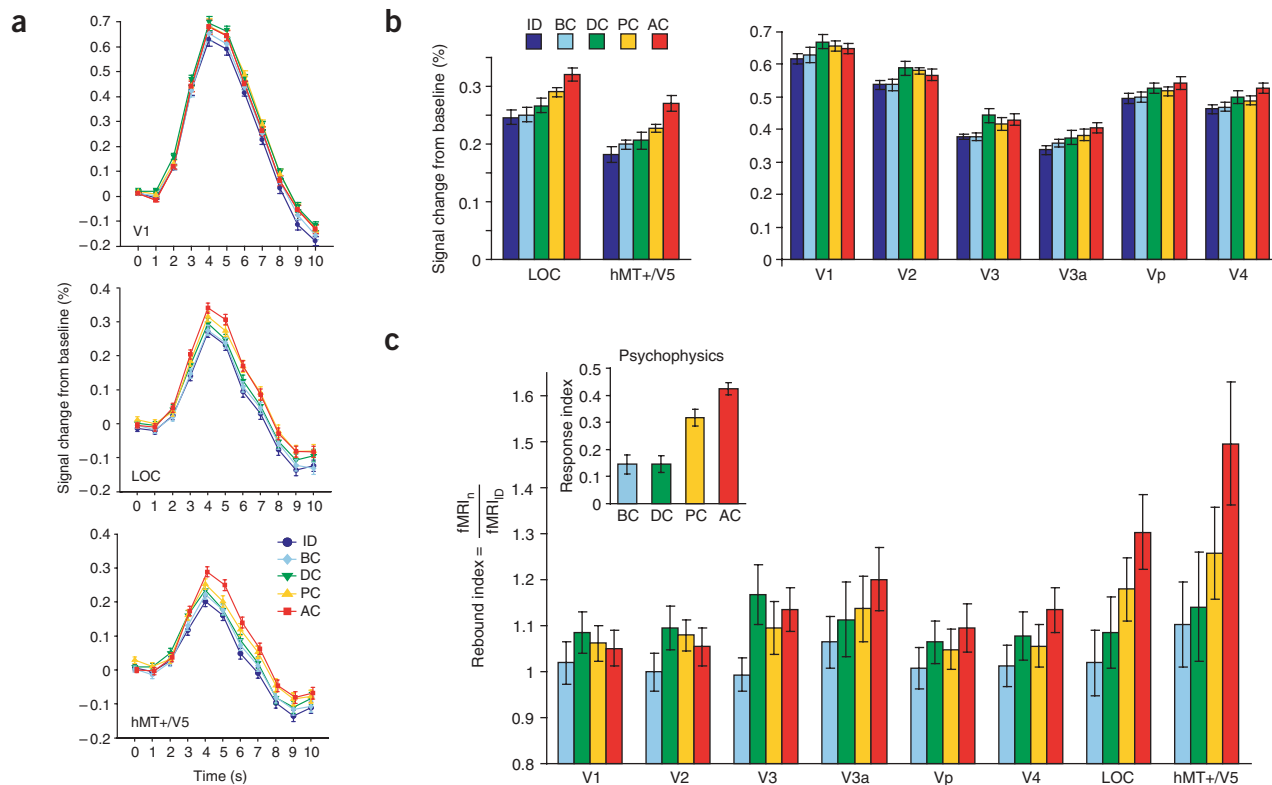


Figure 3 fMRI results from Experiment 1. (a) Representative average time course ($n = 11$) of the fMRI response (areas V1, LOC, hMT+/V5). Error bars, \pm s.e.m. (b) Averaged fMRI response from the peak time points of the fMRI time course in LOC, hMT+/V5 and retinotopic areas (hemispheres combined). Error bars, \pm s.e.m. No differences were found between sub-regions of the LOC¹² or hMT+/V5. (c) fMRI responses expressed as a rebound index across ROIs. Rebound index equals peak fMRI response in a condition (fMRI_n) divided by the response to the Identical Condition (fMRI_{ID}); 1 indicates an adapted response, and values > 1 indicate sensitivity to changes in the stimulus. Error bars (normalized s.e.m.) incorporate the error estimates of both numerator and denominator. Inset: between-subjects mean psychophysical data ($n = 11$) for each experimental condition expressed as a psychophysical response index: 0 represents random behavior, and 0.5 represents perfect discrimination. Error bars, \pm s.e.m.

fMRI responses in areas LOC and hMT+/V5, but not in the early visual areas (Table 1b). Furthermore, to ensure that this effect was not specific to the chosen inconsistent-cue test stimulus, we repeated the experiment with a different test stimulus in which the dihedral angles specified by the disparity and perspective cues were swapped ($S_d = 136^\circ$, $S_p = 116^\circ$). We observed the same relationships under these conditions (Fig. 4), providing complementary evidence that responses in higher, rather than early, visual areas reflect perceived shape based on cue combination.

Experiment 2: examining variations in cue weights

An interesting aspect of previous studies examining cue combination is the inter-subject variability in cue weights. This suggests that observers differ in the reliability with which they encode cues, leading to differences in perception^{11,19}. Consistent with previous psychophysical reports, we observed differences in fMRI response in areas LOC and hMT+/V5 that corresponded to between-subjects differences in perceived 3D shape (Supplementary Table 1, Supplementary Fig. 1).

To examine further the relationship in individual observers between fMRI response and the 'combined-cue' shape percept, we studied fMRI responses to stimuli that were chosen based on an individual's cue weights. Rather than ensuring that observers viewed identical stimuli (Experiment 1), we generated stimuli for individuals after psychophysical tests that measured disparity and perspective cue weights. This allowed us to test for fMRI adaptation to perceived shape using stimuli that differed in their component cues but were reported to have similar 3D shape.

Specifically, before collecting fMRI data, we measured each observer's cue weights (w_d , w_p) inside the magnet. Observers ($n = 3$) made judgments about the dihedral angle of an inconsistent-cue test stimulus ($S_d = 146^\circ$; $S_p = 126^\circ$) by comparisons with consistent-cue reference stimuli. Fitting these psychophysical data provided a measure of perceived angle for the test stimulus (S_c) that was used to estimate the disparity and perspective cue weights. With this knowledge, we created a Balanced Change (BC) test stimulus in which the disparity ($S_d = \delta^\circ$) and perspective ($S_p = \pi^\circ$) cues changed in opposite directions so that when combined, perceived 3D shape ($S_c = w_d\delta + w_p\pi$) was very similar to that of a consistent-cue reference stimulus ($S_d = S_p = 126^\circ$). Two additional inconsistent-cue test stimuli were created (Fig. 5a) that corresponded to individual cue differences between the Balanced Change test stimulus and the consistent-cue reference: the Disparity Change stimulus ($S_d = \delta$, $S_p = 126^\circ$) and Perspective Change stimulus ($S_d = 126^\circ$, $S_p = \pi$). The magnitude of angular slant change between the Disparity and Perspective Change test stimuli and the consistent-cue reference was different (that is, $\Delta d > \Delta p$). However, psychophysical performance was predicted to be similar in these conditions, as the magnitude of cue change reflected each observer's cue weights ($w_d < w_p$). Before collecting the fMRI data, we validated these predictions about the perceived shape of the three test stimuli (Balanced Change, Disparity Change,

Perspective Change) through further psychophysical testing, by comparison with consistent-cue stimuli, as above.

For the fMRI experiment, these tailored test stimuli were used in an event-related adaptation procedure similar to Experiment 1. Four experimental conditions were tested. The Identical Condition, which consisted of a consistent-cue reference stimulus ($S_p = S_d = 126^\circ$) shown twice, provided a baseline measure. The Balanced Change, Disparity Change and Perspective Change conditions consisted of the reference stimuli ($S_d = S_p = 126^\circ$) paired with each of the tailored inconsistent-cue test stimuli (Fig. 5a).

We reasoned that areas engaged in 3D shape perception on the basis of combined cues should show fMRI responses in the Balanced Change condition similar to those in the Identical condition, owing to the perceptual similarity of the test and reference 3D shapes. However, these areas should show increased responses when the cue changes required for the Balanced Change stimulus are made in isolation (Disparity and Perspective Change conditions) because of differences in perceived 3D shape. Furthermore, because the Disparity and Perspective Change conditions had similar perceptual discriminability, similar fMRI responses should be observed in these conditions. fMRI responses in the higher, but not early, visual areas were consistent with these predictions (Fig. 5b–d). Specifically, fMRI responses in LOC and hMT+/V5 for the Balanced Change condition were similar to those for the Identical Condition (Table 2a). In contrast, isolated cue changes (Disparity and Perspective Change conditions) resulted in similar fMRI responses that were significantly higher than responses for the Balanced Change condition, in which the same cue changes were made concurrently. However, fMRI responses in the early visual areas for the Balanced Change and the individual cue change conditions were

Table 1 Statistical analyses for Experiment 1

(a) Statistical analyses (repeated measures ANOVA with Greenhouse-Geisser contrasts) for Experiment 1

ROI	IC: DC		IC: PC		BC+DC: AC+PC		DC: BC	
	$F_{1,280}$	P	$F_{1,280}$	P	$F_{3,280}$	P	$F_{1,280}$	P
V1	9.39	0.014	4.90	0.030	0.40	0.388	5.53	0.027
V2	8.85	0.015	6.17	0.024	0.77	0.382	8.93	0.015
V3	13.78	0.007	4.42	0.033	1.15	0.160	14.86	0.008
V3a	4.96	0.030	7.21	0.020	4.91	0.049	0.93	0.176
Vp	4.29	0.034	2.31	0.110	3.71	0.078	2.97	0.047
V4	3.42	0.042	1.91	0.124	2.12	0.116	2.62	0.049
LOC	1.49	0.141	6.63	0.022	14.70	0.013	0.89	0.180
hMT+/V5	2.21	0.137	7.34	0.019	14.45	0.013	0.17	0.298

(b) Correlation coefficient and associated ANOVAs for regressions on behavioral and fMRI data for each ROI in Experiment 1

ROI	R	$F_{1,42}$	P
V1	0.01	0.02	0.90
V2	0.06	0.16	0.68
V3	0.01	0.01	0.98
V3a	0.12	1.04	0.31
Vp	0.14	0.89	0.35
V4	0.09	0.36	0.54
LOC	0.44	15.87	0.001
hMT+/V5	0.40	12.82	0.001

ROI, region of interest. IC, Identical Condition. DC, Disparity Change. PC, Perspective Change. BC, Balanced Change. AC, All Change.

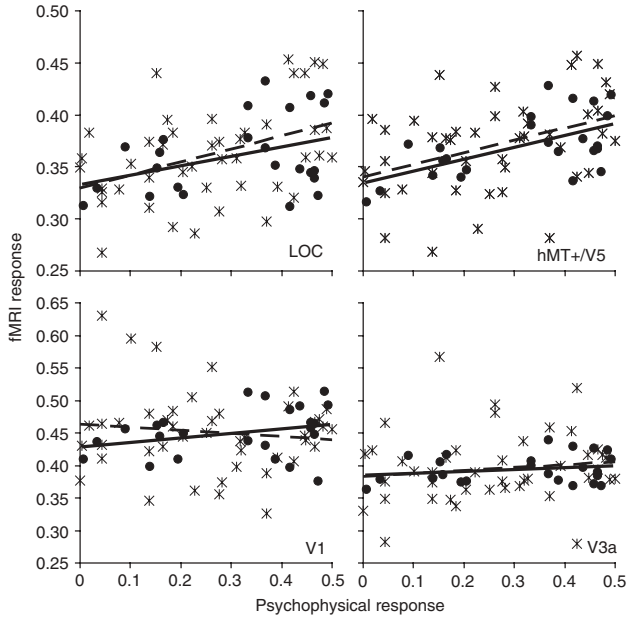


Figure 4 Regressions of fMRI responses on psychophysical responses: fMRI data from Experiment 1 (asterisks, dashed lines) against corresponding psychophysical response index. fMRI data were normalized across subjects ($n = 11$) and areas. Filled circles and solid lines indicate regressions on data obtained from an additional experiment in which a different test stimulus was used (six subjects). We did not observe a difference between the β (slope) coefficient of the regressions for each data set based on the 95% confidence intervals.

not significantly different (**Fig. 5d**, **Table 2a**). Consistent with Experiment 1, these data suggest that fMRI responses in the higher, but not early, visual areas relate to the perception of 3D shape on the basis of cue combination.

Controls for possible confounds

To ensure that our experimental results could not be confounded by differences in eye movement or attentional demands across

conditions, we conducted control experiments and additional analyses. First, eye position measurements indicated that eye movements were very small and were not systematically different across experimental conditions (**Supplementary Fig. 2**). Second, to help maintain correct eye vergence, we used nonius fixation targets for the first (**Supplementary Fig. 2**) and second experiments. Furthermore, in our setup, changes in horizontal disparity would be the main stimulus that could drive vergence eye movements, as perspective cue signals are counteracted by conflicting disparities²⁰. We observed large changes in the fMRI response that were associated with no change in the horizontal disparity signal (Perspective Change conditions), making it unlikely that vergence changes confounded our results. However, to rule out a contribution of eye vergence, we investigated fMRI responses when observers were required to fixate targets across the range of disparities used in our experiments. Using the fMRI design employed in the main experiments, we did not observe significant sensitivity to the small changes in vergence that would be evoked by these disparity changes in any of the measured ROIs (**Fig. 6a**; **Table 2b**). Therefore, it is unlikely that our results were confounded by vergence differences across conditions.

It is also unlikely that differences in allocation of attention could account for the observed pattern of fMRI responses. In contrast to our fMRI results, an attentional load explanation would predict higher

Figure 5 Design and results from Experiment 2. **(a)** Design of Experiment 2, in which inconsistent-cue test stimuli (filled circles) reflected each subject's cue weights (w_d , w_p). The inconsistent-cue stimulus for the Balanced Change (BC) condition was constructed so that perceived shape was very similar to the consistent-cue reference (filled square, $S_d = S_p = 126^\circ$). Values for the BC test stimulus for each subject: S1, $S_d = 96^\circ$, $S_p = 134^\circ$, measured $S_c = 124.4 \pm 2.5^\circ$; S2, $S_d = 96^\circ$, $S_p = 148^\circ$, measured $S_c = 127.3 \pm 2.2^\circ$; S3, $S_d = 98^\circ$, $S_p = 136^\circ$, measured $S_c = 128.5 \pm 1.9^\circ$. The inconsistent-cue stimuli for the Disparity Change (DC) and Perspective Change (PC) conditions consisted of the individual cue changes required for the Balanced Change inconsistent-cue stimulus. **(b)** Representative average time course ($n = 3$) of the fMRI response (areas V1, LOC, hMT+/V5). Error bars, \pm s.e.m. **(c)** Averaged peak fMRI response for each of the three observers in LOC and hMT+/V5. Error bars, \pm s.e.m. **(d)** Averaged peak fMRI responses in LOC, hMT+/V5 and retinotopic visual areas. Between-subjects data presented for illustration ($n = 3$). The same pattern was observed in each observer's data. Error bars, \pm s.e.m.

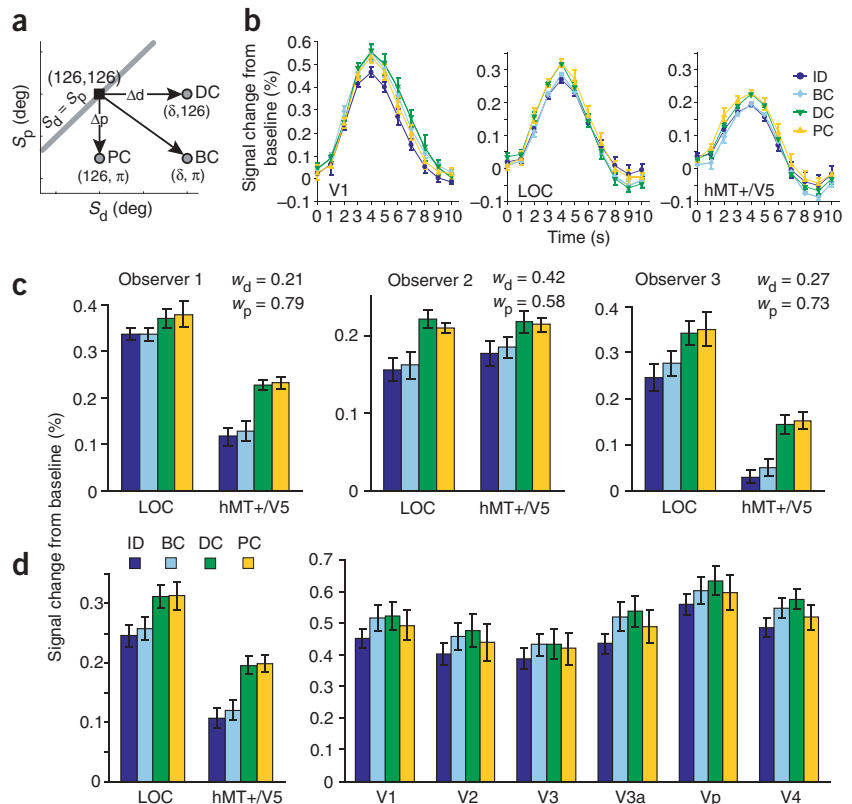


Table 2 Statistical analyses for Experiment 2 and vergence control experiment

(a) Statistical analyses (repeated measures ANOVA with Greenhouse-Geisser contrasts) for Experiment 2.

ROI	DC+PC: BC		ID: BC	
	$F_{1,45}$	P	$F_{1,45}$	P
V1	1.09	0.302	3.37	0.039
V2	1.42	0.240	2.22	0.074
V3	1.51	0.226	1.91	0.085
V3a	1.69	0.201	4.95	0.017
Vp	0.72	0.359	2.10	0.079
V4	1.10	0.300	3.09	0.045
LOC	10.88	0.006	0.18	0.577
hMT+/V5	6.69	0.013	0.50	0.404

(b) Statistical analyses (repeated measures ANOVA) on data obtained from the vergence control experiment.

ROI	All conditions	
	$F_{3,12}$	P
V1	0.035	0.991
V2	0.098	0.959
V3	0.138	0.935
V3a	0.071	0.974
Vp	0.185	0.904
V4	0.055	0.982
LOC	0.360	0.783
hMT+/V5	0.211	0.886

fMRI responses when the discrimination task was hardest, as difficult conditions require prolonged, focused attention, resulting in higher fMRI responses²¹. We observed the opposite effect: fMRI responses were highest in the All Change condition, in which the discrimination was easiest (Fig. 3c: highest mean response index) and subjects responded fastest (Fig. 6b: shortest response times). Further, it is unlikely that observers chose to attend selectively to particular conditions, as trials were presented in quick succession and were randomly interleaved. Finally, it could be argued that perceived change is more interesting than perceived lack of change and, as a result, responses in the high-discriminability conditions were increased. However, such increases in general alertness or arousal would result in increases in fMRI response across the visual areas²¹. This is not consistent with the different patterns of fMRI responses that we observed in the early and higher visual areas.

DISCUSSION

Our study provides neuroimaging evidence for depth cue combination in the human visual cortex and demonstrates the following main findings. First, higher visual areas in both the ventral (LOC) and the dorsal (hMT+/V5) pathways appear to represent the perceived global 3D shape based on a combination of cues to slant. Second, fMRI responses in most of the early visual areas indicate sensitivity to changes in the depth cues that define 3D structure rather than to changes in the 'combined-cue' shape percept. That is, we observed significant fMRI responses to disparity changes across early visual areas and to perspective changes in V1, V2 and dorsal (V3, V3a) extrastriate areas. This observed sensitivity to changes in disparity in early retinotopic cortex is consistent with previous physiological^{22–24} and imaging^{5,18} studies. Our experiments were not designed to delineate the precise nature of this sensitivity. However, it is likely that the responses we have measured reflect a functional hierarchy of disparity processing, ranging from the processing of local disparities at early stages²⁵ to surface-based slant processing at later stages. Previous work investigating the processing of perspective information has focused on parietal areas^{7,26,27}. Our findings raise the possibility that perspective cue processing involves a dorsal network ascending into parietal cortex. Finally, previous studies^{5,6,16,18,28} have implicated V3a in processing 3D information. Our findings suggest a possible role of area V3a in the combination of cues for global 3D shape (Fig. 3). However, the activity in area V3a did not seem to be as tightly coupled to perception as in higher areas (Figs. 4 and 5).

Cue combination: the problem and the process

To date, almost all that we know about the mechanisms of visual cue combination comes from theoretical considerations and psychophysical data. Recent studies have suggested that under many circumstances, the brain combines cues using weighted linear combination¹. Under this scheme, individual cues are initially processed largely independently to yield an estimate of an environmental property. Estimates from different depth cues are then combined by weighting each source of information in proportion to its statistical reliability¹⁹.

The neural mechanisms that mediate depth cue processing have been studied extensively. In particular, several studies provide evidence for retinotopic visual areas involved in processing disparity²³, stereoscopic edges²⁹, 3D orientation³⁰ and surfaces^{5,6,28,31}. Furthermore, monkey IT cortex^{17,32–34} and human LOC^{15,35} have been implicated in processing cues for 3D structure. Finally, MT has been shown to respond to disparity-defined surfaces³⁶ even in the absence of motion^{15,37}, to carry information about 3D surface orientation⁴ and to mediate perception of structure from motion^{36,38}.

Previous work has often considered sensitivity to single depth cues. For instance, disparity information is often 'isolated' using random dot stereograms. However, such viewing conditions often create large cue

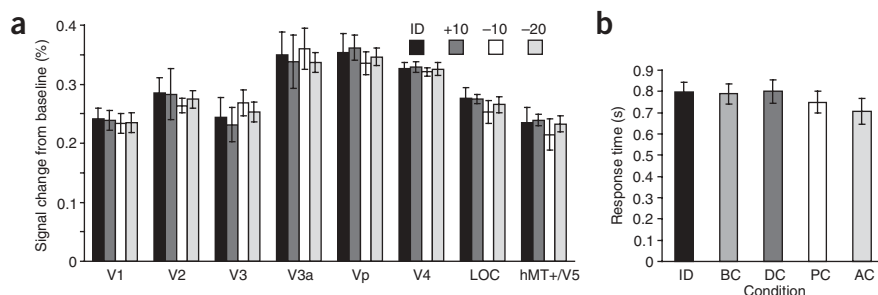


Figure 6 Vergence control experiment and response time data. (a) Vergence control: fMRI responses when observers ($n = 4$) viewed two sequentially presented fixation stimuli at different depth planes representing the range of disparities presented in Experiment 1 (5.5–9.6'). Observers were required to determine whether the first or second viewed stimulus was closer. Error bars, \pm s.e.m. (b) Between-subjects mean response time in each condition (Experiment 1). Error bars, \pm s.e.m. The shortest mean response times were observed in the All Change condition.

conflicts: random dot patterns specify a flat surface (that is, no texture gradient), in contrast to the disparity information. The perturbation approach¹ we have adopted is a useful technique for studying the neural basis of depth cue processing, as it reduces such conflicts. However, to ensure our results were not limited to experimentally induced cue-conflict stimuli (which might exceed natural bounds), we examined the relationship between perceptual discriminability and fMRI responses using consistent-cue stimuli. Consistent with Experiments 1 and 2, we observed stronger fMRI responses for high-discriminability conditions than for low-discriminability conditions in the higher visual areas (**Supplementary Fig. 3**).

Cue combination: neural selectivity and invariance

Our fMRI findings provide insights into the neural correlates of cue combination at the scale of large neural populations and suggest cortical regions in which responses are consistent with the ‘combined-cue’ percept. Further physiological studies (J.D. Nguyenkim & G.C. DeAngelis, *Soc. Neurosci. Abstr.* 368.12, 2004) in these regions, with higher spatiotemporal resolution, are necessary to determine whether the neural responses underlying weighted-cue combination are carried by single neuron responses, population codes or both. These studies could determine whether individual neurons within an area respond on the basis of cue combination, or whether subpopulations within an area carry information about the two cues independently. Our findings provide evidence that even if disparity and perspective information is encoded by separate subpopulations, the response magnitude of these subpopulations in higher areas seems to vary in accordance with ‘combined-cue’ 3D shape perception.

Complementary to feature selectivity, invariance to image changes is a neuronal property important for visual perception and recognition. Recent studies provide evidence for representations independent of the cues that define an object’s shape in monkey IT^{8,32,39} and human LOC^{10,40}. Similarly, neurons in MT have been suggested to compute 3D surfaces from multiple cues^{41,42}. Finally, recent human^{27,43–45} and monkey^{7,16,26,46} fMRI studies suggest a network of occipitotemporal and parietal regions involved in processing 3D objects that are defined by different cues. These studies propose that regions involved in processing different depth cues could also mediate cue combination. However, information about the behaviorally relevant cue weights is necessary to infer cue combination. The perturbation technique we adopted breaks the correlation between cue and percept changes, allowing us to directly test areas involved in cue combination. Our data suggest that occipitotemporal areas previously discussed with reference to cue invariance are implicated in processing combinations of cues. Future studies will address the role of parietal regions, as the high spatiotemporal resolution required for the acquisition of the fMRI adaptation data has not allowed us to measure fMRI responses in these regions.

In conclusion, an exciting challenge in neuroscience is to understand how the same sensory cues can result in different percepts⁴⁷ and how different cues can result in the same percept⁴⁸. Our study provides a first step in understanding neural circuits in human visual cortex that are fundamental for resolving the perceptual ambiguities inherent in combining multiple information sources. Further studies on the neural processing of complex 3D objects⁴⁹ defined by multiple cues within and across modalities¹⁹ will provide important insights into the networks mediating our unified perception of the 3D world.

METHODS

Observers. Observers gave informed written consent. All had normal or corrected-to-normal visual acuity without stereoscopic or color deficits.

Stimuli and psychophysical methods. Stimuli were receding hinged planes ($5 \times 6.6^\circ$) defined by perspective and horizontal disparity (range: 4.3–13.3’) cues to 3D structure. They had the same projection-screen width irrespective of the slant (range: 17–47°) and were surrounded by an irregular background (peripheral extent: 9.8° horizontal, 7.7° vertical) of squares (0.5°) in the plane of the screen, which enhanced relative disparities and was designed to promote accurate eye vergence⁵⁰. Stimuli were rear projected and viewed through a mirror 10 cm above the eyes (viewing distance, 78 cm). Red-green anaglyphs were used (crosstalk was only 0.6%). Photometric measures of the red and green signals from the NECGT950 video projector were used for gamma correction and to equate image luminance for each eye. Stimuli for Experiments 1 (replication; **Supplementary Fig. 2**) and 2 included a pair of nonius lines on each side of the fixation circle to promote stable eye vergence.

To evaluate the perceived shape of the test stimuli observers made comparisons with a range (86–146°) of consistent-cue reference stimuli (method of constant stimuli). Observers fixated a circle (diameter = 0.4°) at the stimulus’ center and judged which of two sequentially presented stimuli (test and reference order randomized) had a larger dihedral angle. Responses were collected using optical response buttons. The PSE was calculated by fitting (Probit analysis) the proportion of ‘larger dihedral angle’ responses to the test stimulus at each consistent-cue reference angle. Before collecting fMRI data we obtained a full psychometric function inside the magnet (e.g. **Fig. 1d**). For Experiment 1, psychophysical measurements made concurrently with fMRI acquisition provided data for four points on the function (filled squares **Fig. 1d**). PSEs obtained with and without the scanner running did not differ ($t_{10} = 2.23$, $P = 0.52$). We calculated a psychophysical response index to describe the data. This is the absolute difference between the ‘larger dihedral angle’ response proportion in a condition and 0.5: 0 represents random behavior and 0.5 perfect discrimination. A criterion of 0.25 was used to distinguish high- and low-discriminability conditions.

Cue weights were calculated by assuming weighted linear combination where weights sum to 1 (ref. 1): $S_c = w_d S_d + w_p S_p$, $w_d + w_p = 1$. When tailoring stimuli to observers’ cue weights (Experiment 2), we calculated stimulus values using these formulas so that perceived 3D shape in the Balanced Change condition would be very similar to the consistent-cue reference (126°). To ensure these calculations were correct, we evaluated each observer’s perception of the inconsistent-cue test stimuli (BC, DC, PC) prior to collecting fMRI data (as above).

Imaging. Data were collected using a 3-T Siemens scanner with gradient echo pulse sequence (TR = 2 s, TE = 40 ms, localizers; TR = 1 s, TE = 40 ms, event-related scans) and a head coil. The high resolution of the event-related scans limited us to 11 axial slices (5 mm thick; 3×3 mm in-plane resolution) covering occipitotemporal regions. Subjects ran one session of nine scans: five localizers (LOC, twice; hMT+/V5, once; retinotopic, twice) and four (Experiment 1) or six (Experiment 2) event-related scans (order counterbalanced across subjects). For event-related experiments, each scan consisted of one experimental trial epoch and two 8-s fixation epochs (one at the start and one at the end). For Experiment 1, each scan had 24 experimental trials per condition ($n = 5$) and 24 fixation trials (providing a measure of baseline activity). For Experiment 2, each scan had 25 experimental trials per condition ($n = 4$) and 25 fixation trials. Presentation order was counterbalanced so that trials from each condition (including fixation) were preceded equally often by trials from other conditions. A new trial began every 3 s and consisted of two stimuli, each presented for 400 ms (ISI = 150 ms), followed by a blank (2,050 ms). Subjects did not perceive apparent motion between the stimuli. Note that changes in disparity and perspective cues result in changes in spatial position. The magnitude of these low-level stimulus changes was deliberately kept very small (largest difference in position of corners was 6’).

Data analysis. fMRI data were processed using BrainVoyager (Brain Innovation). For each individual subject, regions of interest (ROIs: V1, V2, V3, V3a, VP, V4, hMT+/V5, LOC) were defined using standard techniques^{10,15} (**Fig. 2** and **Supplementary Fig. 4**). We averaged the signal intensity across trials in each condition at each time point and converted these to percentage signal change relative to fixation for each ROI and subject¹⁰. Sample between-subject mean fMRI time courses in areas V1, LOC and hMT+/V5 are provided

(Figs. 3a and 5b); time courses in other areas were similar. Fitting the time courses with the hemodynamic response function indicated that peak fMRI responses occurred between 4 and 5 s after trial onset (time-to-peak parameter for every ROI shown in **Supplementary Fig. 5**). Statistical analyses were performed on the average fMRI response at these time points (repeated-measures ANOVAs, Greenhouse-Geisser contrasts).

Finally, we controlled for the possibility that differences between experimental conditions were due to different fMRI responses evoked by the consistent-cue reference stimuli themselves (Experiment 1). We tested fMRI responses when each consistent-cue reference stimulus was presented alone in a trial. There were no significant differences between fMRI responses for the different stimuli in any ROI (**Supplementary Fig. 6**), ensuring that fMRI responses in the adaptation experiments were due to differences between the test and reference stimuli.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

Preliminary reports of this work were presented at the VisionSciences Society's 2003 meeting and at the Society for Neuroscience's 2003 meeting. Thanks to S. Maier and J. Lam for help with data collection, and N. Logothetis, N. Kanwisher, J. Harris, S. McDonald, M. Ernst and R. Fleming for helpful discussions and comments. Thanks also to R. van Ee for advice on stimulus generation. Supported by an Alexander von Humboldt Fellowship to A.E.W., the Max-Planck Society and DFG grant TH812/1-1.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 17 February; accepted 12 April 2005

Published online at <http://www.nature.com/natureneuroscience/>

- Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. Measurement and modeling of depth cue combination – in defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
- Clark, J.J. & Yuille, A.L. *Data fusion for sensory information processing systems* (Kluwer Academic, Boston, 1990).
- Hillis, J.M., Ernst, M.O., Banks, M.S. & Landy, M.S. Combining sensory information: mandatory fusion within, but not between, senses. *Science* **298**, 1627–1630 (2002).
- Nguyenkim, J.D. & DeAngelis, G.C. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J. Neurosci.* **23**, 7117–7128 (2003).
- Backus, B.T., Fleet, D.J., Parker, A.J. & Heeger, D.J. Human cortical activity correlates with stereoscopic depth perception. *J. Neurophysiol.* **86**, 2054–2068 (2001).
- Tsao, D.Y. *et al.* Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron* **39**, 555–568 (2003).
- Tsutsui, K., Sakata, H., Naganuma, T. & Taira, M. Neural correlates for perception of 3D surface orientation from texture gradient. *Science* **298**, 409–412 (2002).
- Liu, Y., Vogels, R. & Orban, G.A. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *J. Neurosci.* **24**, 3795–3800 (2004).
- Grill-Spector, K. & Malach, R. fMRI-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)* **107**, 293–321 (2001).
- Kourtzi, Z. & Kanwisher, N. Representation of perceived object shape by the human lateral occipital complex. *Science* **293**, 1506–1509 (2001).
- Knill, D.C. & Saunders, J.A. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* **43**, 2539–2558 (2003).
- Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. The lateral occipital complex and its role in object recognition. *Vision Res.* **41**, 1409–1422 (2001).
- Zeki, S. *et al.* A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* **11**, 641–649 (1991).
- Tootell, R.B. *et al.* Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
- Kourtzi, Z., Buelthoff, H.H., Erb, M. & Grodd, W. Object-selective responses in the human motion area MT/MST. *Nat. Neurosci.* **5**, 17–18 (2002).
- Orban, G.A. *et al.* Similarities and differences in motion processing between the human and macaque brain: evidence from fMRI. *Neuropsychologia* **41**, 1757–1768 (2003).
- Janssen, P., Vogels, R. & Orban, G.A. Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* **288**, 2054–2056 (2000).
- Neri, P., Bridge, H. & Heeger, D.J. Stereoscopic processing of absolute and relative disparity in human visual cortex. *J. Neurophysiol.* **92**, 1880–1891 (2004).
- Ernst, M.O. & Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
- Enright, J.T. Art and the oculomotor system: perspective illustrations evoke vergence changes. *Perception* **16**, 731–746 (1987).
- Ress, D., Backus, B.T. & Heeger, D.J. Activity in primary visual cortex predicts performance in a visual detection task. *Nat. Neurosci.* **3**, 940–945 (2000).
- Thomas, O.M., Cumming, B.G. & Parker, A.J. A specialization for relative disparity in V2. *Nat. Neurosci.* **5**, 472–478 (2002).
- Cumming, B.G. & Parker, A.J. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* **389**, 280–283 (1997).
- Watanabe, M., Tanaka, H., Uka, T. & Fujita, I. Disparity-selective neurons in area V4 of macaque monkeys. *J. Neurophysiol.* **87**, 1960–1973 (2002).
- Nienborg, H., Bridge, H., Parker, A.J. & Cumming, B.G. Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *J. Neurosci.* **24**, 2065–2076 (2004).
- Taira, M., Tsutsui, K.I., Jiang, M., Yara, K. & Sakata, H. Parietal neurons represent surface orientation from the gradient of binocular disparity. *J. Neurophysiol.* **83**, 3140–3146 (2000).
- Shikata, E. *et al.* Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. *J. Neurophysiol.* **85**, 1309–1314 (2001).
- Poggio, G.F., Gonzalez, F. & Krause, F. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J. Neurosci.* **8**, 4531–4550 (1988).
- von der Heydt, R., Zhou, H. & Friedman, H.S. Representation of stereoscopic edges in monkey visual cortex. *Vision Res.* **40**, 1955–1967 (2000).
- Hinkle, D.A. & Connor, C.E. Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.* **5**, 665–670 (2002).
- Bakin, J.S., Nakayama, K. & Gilbert, C.D. Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neurosci.* **20**, 8188–8198 (2000).
- Tanaka, H., Uka, T., Yoshiyama, K., Kato, M. & Fujita, I. Processing of shape defined by disparity in monkey inferior temporal cortex. *J. Neurophysiol.* **85**, 735–744 (2001).
- Uka, T., Tanaka, H., Yoshiyama, K., Kato, M. & Fujita, I. Disparity selectivity of neurons in monkey inferior temporal cortex. *J. Neurophysiol.* **84**, 120–132 (2000).
- Janssen, P., Vogels, R., Liu, Y. & Orban, G.A. Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *J. Neurosci.* **21**, 9419–9429 (2001).
- Gilaie-Dotan, S., Ullman, S., Kushnir, T. & Malach, R. Shape-selective stereo processing in human object-related visual areas. *Hum. Brain Mapp.* **15**, 67–79 (2002).
- Bradley, D.C., Chang, G.C. & Andersen, R.A. Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* **392**, 714–717 (1998).
- Palanca, B.J. & DeAngelis, G.C. Macaque middle temporal neurons signal depth in the absence of motion. *J. Neurosci.* **23**, 7647–7658 (2003).
- Dodd, J.V., Krug, K., Cumming, B.G. & Parker, A.J. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* **21**, 4809–4821 (2001).
- Sary, G., Vogels, R. & Orban, G.A. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* **260**, 995–997 (1993).
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y. & Malach, R. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21**, 191–202 (1998).
- Albright, T.D. Cortical processing of visual motion. *Rev. Oculomot. Res.* **5**, 177–201 (1993).
- Treue, S. & Andersen, R.A. Neural responses to velocity gradients in macaque cortical area MT. *Vis. Neurosci.* **13**, 797–804 (1996).
- Murray, S.O., Olshausen, B.A. & Woods, D.L. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb. Cortex* **13**, 508–516 (2003).
- Paradis, A.L. *et al.* Visual perception of motion and 3-D structure from motion: an fMRI study. *Cereb. Cortex* **10**, 772–783 (2000).
- Taira, M., Nose, I., Inoue, K. & Tsutsui, K. Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. *Neuroimage* **14**, 959–966 (2001).
- Sereno, M.E., Trinath, T., Augath, M. & Logothetis, N.K. Three-dimensional shape representation in monkey cortex. *Neuron* **33**, 635–652 (2002).
- van Ee, R., van Dam, L.C. & Erkelens, C.J. Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants. *J. Vis.* **2**, 597–607 (2002).
- Backus, B.T. in *Advances in Neural Information Processing Systems 14* (eds Dietterich, T.G., Becker, S. & Ghahramani, Z.) (MIT Press, Cambridge, Massachusetts, 2002).
- Todd, J.T. The visual perception of 3D shape. *Trends Cogn. Sci.* **8**, 115–121 (2004).
- DeAngelis, G.C. & Newsome, W.T. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* **19**, 1398–1415 (1999).