nature neuroscience

The integration of motion and disparity cues to depth in dorsal visual cortex

Hiroshi Ban[1,2], Tim J Preston[1,3], Alan Meeson[1] & Andrew E Welchman[1]

Humans exploit a range of visual depth cues to estimate three-dimensional structure. For example, the slant of a nearby tabletop can be judged by combining information from binocular disparity, texture and perspective. Behavioral tests show humans combine cues near-optimally, a feat that could depend on discriminating the outputs from cue-specific mechanisms or on fusing signals into a common representation. Although fusion is computationally attractive, it poses a substantial challenge, requiring the integration of quantitatively different signals. We used functional magnetic resonance imaging (fMRI) to provide evidence that dorsal visual area V3B/KO meets this challenge. Specifically, we found that fMRI responses are more discriminable when two cues (binocular disparity and relative motion) concurrently signal depth, and that information provided by one cue is diagnostic of depth indicated by the other. This suggests a cortical node important when perceiving depth, and highlights computations based on fusion in the dorsal stream.

To achieve robust estimates of depth, the brain combines information from different visual cues[1–3]. Computational work proposes this produces more reliable estimates[4] and behavioral tests show it makes objects easier to discern[5,6]. However, our understanding of the neural basis of integration is underdeveloped. Electrophysiological recordings suggest locations where depth signals converge[7–9]. Nevertheless, comparing the responses evoked by individual cues presented 'alone' (for example, disparity-, perspective- or motion-defined depth) does not imply fusion: response characteristics might be dominated by one cue or might show opposite tuning rather than integration[10,11].

Here we used human fMRI to test for cortical areas that integrate cues, rather than containing convergent information (that is, colocated, independent signals). To this end, we exploited two cues to which the brain is particularly sensitive: horizontal binocular disparity and depth from relative motion[12]. Psychophysical evidence for interactions between them[13–16] suggests common stages of processing; thus, these cues provide a useful pairing to test fusion.

To frame the problem of cue integration, consider a solid object (for example, a ballerina) whose depth is defined by both disparity and motion (**Fig. 1a**). An estimate of depth could be derived from each cue (quasi-)independently, defining a bivariate likelihood estimate in motion-disparity space. Thereafter, a fusion mechanism would produce a univariate 'depth' estimate with lower variance[3,4]. To probe this process, it is customary to measure discrimination performance; for instance, asking observers to judge which of two shapes has greater depth (**Fig. 1b**). There are two computationally distinct ways of solving this task: independence or fusion. Under independence, an ideal observer would discriminate the two bivariate distributions (**Fig. 1b**) orthogonal to the optimal decision boundary. By so doing, the observer will be more sensitive to differences between the shapes than if they judge only one cue. This improvement corresponds to the quadratic sum of the discriminabilities of the marginal distributions and has an intuitive geometrical interpretation: by the Pythagorean theorem, the separation between shapes is greater along the diagonal than along the component dimensions.

The alternative possibility is an optimal fusion mechanism that combines the component dimensions into a single ('depth') dimension. This reduces variance, thereby improving shape discrimination. Disparity and motion typically signal the same structure, making the predictions of independence and fusion equivalent (**Fig. 1b**). However, the alternatives can be dissociated by manipulating the viewed shapes experimentally (**Fig. 1c**,**d**), to effect different predictions for independence (**Fig. 1e**) and fusion (**Fig. 1f**).

Here we tested for cue integration at the levels of behavior and fMRI responses. We presented a central plane that was nearer or farther than its surround (**Fig. 2a**). When viewing this stimulus, some neurons will respond to near positions and others far[17], producing a dissociable pattern of activity. fMRI measures this activity at the scale of neuronal populations; nevertheless, multivoxel pattern analysis (MVPA) provides a sensitive tool to reveal depth selectivity in human cortex[18]. Here we decoded fMRI responses evoked when viewing stimuli that depicted near or far depths defined by binocular disparity, relative motion and these signals in combination.

We developed three tests for integration. First, we assessed whether discrimination performance in combined cue settings exceeds quadratic summation. Our logic was that a fusion mechanism is compromised when 'single' cues are presented (**Fig. 1c**). For example, a 'single cue' disparity stimulus contains motion information that the viewed surface is flat, depressing performance (contrast single cues in **Fig. 1e** versus **1f**). Thus, if 'single cue' data are used to derive a prediction for the concurrent stimulus, measured performance will exceed quadratic summation. We used this test to establish a minimum bound for fusion, as considerations of fMRI signal generation and measurement (for example, scanner noise) entail that this test cannot rule out

[1]School of Psychology, University of Birmingham, Edgbaston, Birmingham, UK. [2]Japan Society for the Promotion of Science, Tokyo, Japan. [3]Department of Psychology, University of California, Santa Barbara, Santa Barbara, California, USA. Correspondence should be addressed to A.E.W. (a.e.welchman@bham.ac.uk).

Received 24 October 2011; accepted 10 January 2012; published online 12 February 2012; doi:10.1038/nn.3046

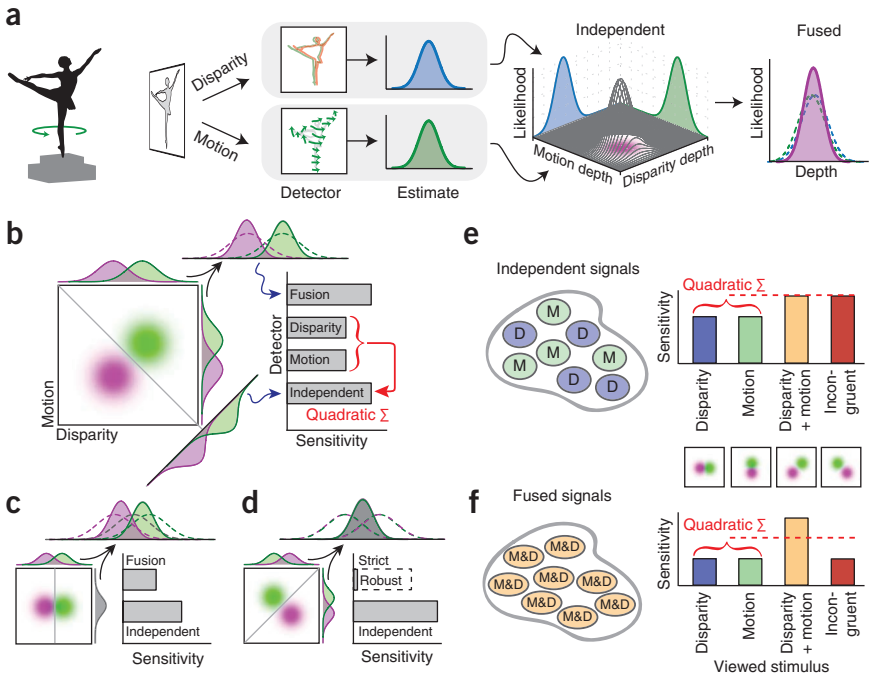

Figure 1 Schematic illustrations of cue fusion and ideal observer discrimination. (**a**) Cartoon of depth processing: depth of the ballerina figurine is estimated from disparity and motion, producing a bivariate Gaussian (purple blob in three-dimensional plot). Fusion combines disparity and motion using maximum likelihood estimation, producing a univariate depth estimate. (**b**) Discriminating two shapes defined by bivariate Gaussians (purple and green blobs). We envisage four types of detector: the disparity type and motion type respond to only one dimension (that is, discrimination of the marginals); the independent detector (bottom right) uses the optimal separating plane (gray line on the diagonal); the fusion detector (top right) integrates cues. (**c**) 'single cue' case: shapes differ in disparity but motion is the same. The optimal separating plane is now vertical (independent detector), and the fusion mechanism is compromised. (**d**) Incongruent cues: disparity and motion indicate opposite depths. Independent performance matches **b**; fusion is illustrated for two scenarios: strict (detector is insensitive) and robust (bar with dashed outline: performance reverts to that of one component). (**e**) Predicted measurements of independent units. Four types of stimuli are displayed: disparity (as in **c**), motion (motion indicates a depth difference, disparity specifies the same depth), disparity + motion (as in **b**), and incongruent (as in **d**). (**f**) Predicted measurements of fused units. Note that performance in the motion and the disparity conditions is lower than in **e**.

independence (see Discussion). Second, we determined whether improved performance is specific to congruent cues (**Fig. 1e** versus **1f**). An independence mechanism should be unaffected by incongruency (**Fig. 1d**), as quadratic summation ignores the sign of differences. However, a fusion mechanism would be affected: a strict fusion mechanism would be insensitive, whereas a robust mechanism would revert to a single component. Third, motivated by psychophysical reports of cross-adaptation between cues[13–15], we determined whether depth from one cue (for example, disparity) is diagnostic of depth from the other (for example, motion).

We found that fMRI responses from the V3B and kinetic occipital (KO) region (which we denote as area V3B/KO) supported stimulus decoding that surpassed the minimum bound, was specific for consistent depth cues and supported a transfer between cues. This suggests a region involved in representing depth from integrated cues, whose activity may underlie improved behavioral performance in multi-cue settings.

RESULTS

Psychophysics

We presented participants with random dot patterns (**Fig. 2b**) depicting depth from (i) binocular disparity, (ii) relative motion and (iii) the combination of disparity and motion. To test for integration psychophysically, we presented two stimuli sequentially with a slight depth difference between them and participants decided which had the greater depth (that is, which was nearer for near stimuli, or farther for far stimuli). Using a staircase procedure, we assessed observers' sensitivity under four conditions by measuring just noticeable difference (j.n.d.) thresholds (**Fig. 2c**). We found that observers were most sensitive when disparity and motion concurrently signaled depth differences, and least sensitive for motion-defined differences. Using performance in the 'single cue' (disparity alone or motion alone) conditions, we generated a quadratic summation prediction for the combined cue (disparity and motion) case. In line with the expectations of fusion, performance for congruent cues exceeded quadratic summation ($F_{1,6} = 8.16$; $P = 0.015$). Moreover, when disparity and motion were incongruent, sensitivity was lower ($F_{1,6} = 11.07$; $P = 0.016$)

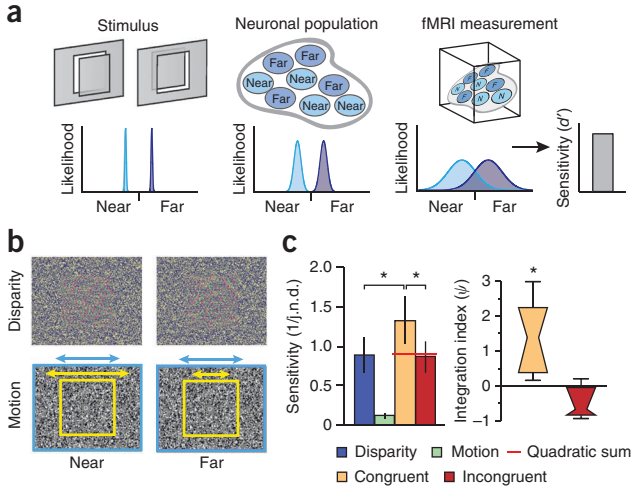

Figure 2 Stimulus illustration and psychophysical results. (**a**) Cartoon of the decoding approach. Participants view stimuli that depict near or far depths. These differentially excite neuronal populations in an area of cortex. fMRI measurements reduce the resolution. We characterize the sensitivity of a decoding algorithm in discriminating near and far stimuli. (**b**) Disparity-defined and motion-defined depth stimuli. The top row provides a red-green anaglyph stereogram. The bottom row provides a cartoon of the relative motion stimuli: yellow arrow, speed of target; blue arrow, speed of background. (**c**) Behavioral tests of integration. Left, observers' mean sensitivity ($N = 7$) with between-subjects s.e.m. Red horizontal line indicates the quadratic summation prediction. Right, results as an integration index for the congruent and incongruent conditions. A value of zero indicates the minimum bound for fusion. Data are presented as notched distribution plots. The center of the 'bowtie' represents the median, the ends depict 68% confidence values, and the upper and lower error bars 95% confidence intervals. $^*P < 0.05$.

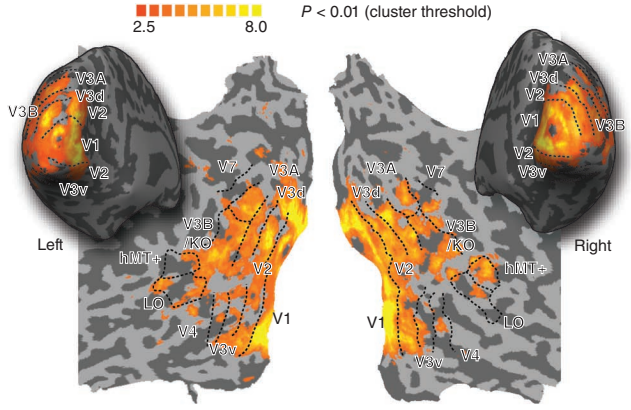

Figure 3 Representative flat maps showing the left and right visual regions of interest from one participant. The maps show the location of retinotopic areas, V3B/KO, the human motion complex (hMT+) and the lateral occipital (LO) area. Regions were defined using independent localizers. Sulci are coded in darker gray than the gyri. Superimposed on the maps are the results of a group searchlight classifier analysis that moved iteratively throughout the entire volume of cortex measured, discriminating between near and far depth positions[18]. The color code represents the *t*-value of the classification accuracies obtained. This analysis confirmed that we had not missed any important areas outside those localized independently.

and comparable to performance in the 'single cue' disparity condition ($F_{1,6} < 1$; $P = 0.809$). To quantify this effect, we calculated a psychophysical integration index (ψ):

$$\psi = \frac{S_{D+M}}{\sqrt{S_D^2 + S_M^2}} - 1 \quad (1)$$

where S_{D+M} is the observer's sensitivity (1/j.n.d.) in the combined condition, and S_D and S_M correspond to sensitivity in the 'single cue' conditions (see ref. 19). A value of zero indicates the minimum bound for fusion (that is, the quadratic sum). Bootstrapping the index revealed that observers' sensitivity exceeded the minimum bound for consistent ($P < 0.001$) but not inconsistent ($P = 0.865$) cue conditions. Additional tests (**Supplementary Fig. 1**) provided further psychophysical evidence of cue integration.

fMRI quadratic summation

To examine the neural basis of disparity and motion integration, we measured fMRI responses in independently localized regions of interest (**Fig. 3**). We then used MVPA to determine which areas contained fMRI signals that enabled a machine learning classifier (support vector machine; SVM) to discriminate reliably between targets presented closer or farther than the fixation plane.

Both disparity- and motion-defined depth were decoded reliably by the classifier, and there was a clear interaction between conditions and areas (**Fig. 4a**; $F_{7.1,135.1} = 6.50$, $P < 0.001$). However, our principal interest was not in 'single cue' processing, or in contrasting overall prediction accuracies between areas (these are influenced by a range of non-neuronal factors). Rather, we were interested in relative performance under conditions in which disparity and motion concurrently signaled depth. Prediction accuracies for the concurrent stimulus were statistically higher than the component cue accuracies in areas V3A ($F_{2,38} = 7.07$; $P = 0.002$) and V3B/KO ($F_{1.5,28.9} = 14.35$; $P < 0.001$). To assess integration, we calculated the minimum bound prediction (**Fig. 4a**) based on quadratic summation. We found that fMRI responses in V3B/KO supported decoding performance that exceeded the minimum bound ($F_{1,19} = 4.99$, $P = 0.019$), but not elsewhere. We quantified this effect across areas using an fMRI integration index (ϕ):

$$\phi = \frac{d'_{D+M}}{\sqrt{d'^2_D + d'^2_M}} - 1 \quad (2)$$

where d'_{D+M} is the classifier's performance in the congruent condition, and d'_D and d'_M are performance for 'single cue' conditions. The values of ϕ differed between areas (**Fig. 4b**; $F_{4.5,86.6} = 3.14$, $P = 0.014$), with a value significantly above zero only in V3B/KO (**Table 1**). This suggests an area in which improved decoding performance may result from the fusion of disparity and motion (although this test cannot rule out independence).

A possible concern is that there may be a gain change in the fMRI response when testing disparity and motion concurrently relative to single cues, and this may enhance the classifier's decoding accuracy (for example, in V3B/KO). However, fMRI signals in each region of interest (**Supplementary Fig. 2a**) showed no evidence for reliable differences in responsiveness between conditions ($F_{2,38} = 2.51$, $P = 0.094$). Another possibility is that fMRI noise may be reduced when cues concurrently signal depth, supporting better decoding. To assess this possibility, we created a composite data set by averaging raw fMRI responses from the 'single cue' conditions. However, prediction accuracies were lower for this composite data set than for the concurrent condition in V3B/KO, indicating that a simple noise reduction did not explain the result (**Supplementary Fig. 2b**; $F_{4.9,93.8} = 3.74$, $P = 0.004$).

Congruent versus incongruent cues

To provide a stronger test for integration, we manipulated both disparity and motion, but placed these cues in extreme conflict (that is, an exaggerated conflict over our 'single cue' conditions). For each stimulus, one cue signaled 'near' and the other 'far' (**Fig. 1d**). If depth from the two cues is independent, this manipulation should have no

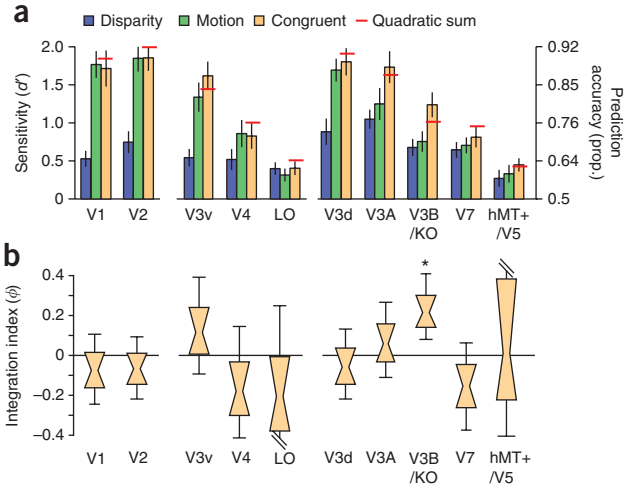

Figure 4 Results for the quadratic summation test. (**a**) Prediction performance for near versus far discrimination in different regions of interest. The red lines illustrate performance expected from the quadratic summation of prediction sensitivities for the 'single cue' conditions. Error bars, s.e.m.; prop., proportion. (**b**) Results as an integration index. A value of zero indicates the minimum bound for fusion (that is, the prediction based on quadratic summation). Data are presented as notched distribution plots. The center of the 'bowtie' represents the median, the shaded area depicts 68% confidence values, and the upper and lower error bars 95% confidence intervals.

Table 1 Significance tests for the integration index (ϕ) and congruency results

	P-value	
Cortical area	Integration index above zero	Congruent versus incongruent
V1	0.789	0.523
V2	0.799	0.419
V3v	0.150	0.079
V4	0.880	0.486
LO	0.838	0.262
V3d	0.733	0.203
V3A	0.265	0.148
V3B/KO	0.001	0.004
V7	0.915	0.247
hMT+/V5	0.479	0.499

Probabilities associated with obtaining a value of zero for (i) the fMRI integration index, and (ii) the prediction accuracy difference between congruent and incongruent stimulus conditions. Values are from a bootstrapped resampling of the individual participants' data. Underlining indicates Bonferroni-corrected significance ($P < 0.05$). LO, lateral occipital cortex; hMT+, human motion complex.

effect. (Note that the classifier distinguishes the stimulus classes that evoked voxel responses; thus, an objectively correct answer exists for the learning algorithm).

Consistent with the idea that V3B/KO fuses signals, discrimination performance was significantly lower when motion and disparity conflicted (**Fig. 5a** and **Table 1**), with accuracy falling to that seen with the 'single cue' components. There was a significant difference between congruent and incongruent conditions ($F_{1,6} = 7.49$, $P = 0.034$), but no significant difference between the incongruent condition and the 'single cue' disparity ($F_{1,6} < 1$, $P = 0.62$) or relative motion ($F_{1,6} = 1.13$, $P = 0.33$) conditions. This robust behavior in the face of extreme conflicts matches perception: conflicts are accommodated within bounds, but thereafter one component is ignored[20]. Our participants relied on disparity when perceiving the incongruent stimulus (**Fig. 2c**). Other visual areas (notably V3v, V3d and V3A), also supported lower prediction accuracies for the incongruent cues (**Fig. 5a**), although these differences were not statistically reliable (**Table 1**).

Transfer test

To obtain a further test for similarities in responses to the two cues, we asked whether depth information provided by one cue (for example, disparity) is diagnostic of depth indicated by the other (for example, motion). We performed a cross-cue transfer test whereby we trained a machine learning classifier to discriminate depth configurations using one cue, and tested the classifier's predictions for data obtained when depth was indicated by the other cue.

To accompany this analysis, we used a control condition that addressed differences in average velocity that arose from the relative motion stimuli. In particular, when we presented motion-defined depth, the classifier might have discriminated movement speed rather than depth position (this likely explains high accuracies for motion in early visual areas; **Fig. 4a**). To control for speed differences, we presented stimuli in which the central target region moved with a fast or slow velocity but there was no moving background, meaning that participants had no impression of relative depth. We reasoned that an area showing a response specific to depth would support transfer between relative motion and disparity, but not between the motion control and disparity.

We observed a significant interaction between accuracy in the transfer tests across regions of interest (**Fig. 5b**; $F_{9,63} = 3.88$, $P = 0.001$). In particular, higher responses for the depth transfer (disparity–relative motion) than the control (disparity–control) were significant in areas V4, V3d and V3B/KO (**Table 2**). To assess the relationship between transfer classification performance (d'_T) and the mean performance for the component cues (that is, $(d'_D + d'_M)/2$), we calculated a bootstrapped transfer index,

$$T = \frac{2d'_T}{d'_D + d'_M} \quad (3)$$

This suggested that transfer test performance was most similar to within-cue decoding in area V3B/KO (**Fig. 5c**). Specifically, transfer performance was around 80% of that obtained when training and testing on the same stimuli. To assess the amount of transfer that arises by chance, we conducted the transfer test on randomly permuted data (1,000 tests per area). This baseline value (see **Fig. 5c**) indicated that transfer between cues was significant in areas V3d and V3B/KO (**Table 2**). In conjunction with the results presented above, this suggests that responses in V3B/KO relate to a more generic representation of depth.

Decoding simulated populations

So far, we have considered two extreme scenarios: independence versus fusion. However, there are computational and empirical reasons to believe that responses might lie between these poles. Computationally, it is attractive to estimate depth based on both fusion and independence,

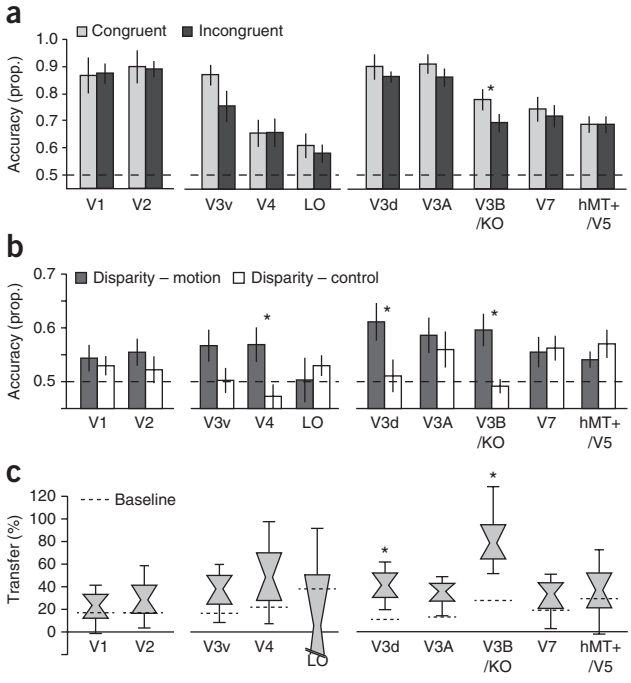

Figure 5 Results for tests of congruency and transfer between cues. (**a**) Prediction accuracy for near versus far classification when cues are congruent (**Fig. 1b**) or incongruent (**Fig. 1d**). The dashed horizontal line at 0.5 corresponds to chance performance for this binary classification. (**b**) Prediction accuracy for the cross-cue transfer analysis. Two types of transfer are depicted: between motion and disparity (gray bars) and between disparity and a flat motion control stimulus (white bars). Classification accuracies are generally lower than for the standard SVM analysis (**Fig. 4a**); this is as expected given the considerable differences between the stimuli that evoked the training and test fMRI responses. Error bars, s.e.m.; prop., proportion. (**c**) Data shown as a transfer index. A value of 100% would indicate that prediction accuracies were equivalent for within- and between- cue testing. Distribution plots show the median, 68% and 95% confidence intervals. Dotted horizontal lines depict a bootstrapped chance baseline based on the upper 95th percentile for transfer obtained with randomly permuted data. $^*P < 0.05$.

Table 2 Significance tests for the between-cue transfer results

Cortical area	*P*-value: Difference between transfer and control accuracies	*P*-value: Transfer index from chance
V1	0.273	0.279
V2	0.068	0.168
V3v	0.024	0.061
V4	0.002	0.102
LO	0.778	0.758
V3d	0.001	0.002
V3A	0.121	0.012
V3B/KO	<0.001	<0.001
V7	0.590	0.141
hMT+/V5	0.815	0.302

Probabilities associated with obtaining zero difference between (i) decoding performance in the disparity-to-relative motion and disparity-to-motion control transfer tests, and (ii) the value of the transfer index in the disparity-to-relative motion condition compared to random (shuffled) performance. These *P*-values are calculated using bootstrapped resampling with 10,000 samples. Underlining indicates Bonferroni-corrected significance ($P < 0.05$). LO, lateral occipital cortex; hMT+, human motion complex.

to determine whether or not cues should be integrated[21]. Empirically, it is unlikely we sampled voxels that respond only to fused signals, as our region of interest localizers were standardized tests that do not target fusion. Thus, it is probable that some voxels (even within V3B/KO) do not reflect integrated cues. To evaluate how a population mixture might affect decoding results, we used simulations to vary systematically the composition of the neuronal population. We decoded simulated voxels whose activity reflected neural maps on the basis of (i) fused depth, (ii) interdigitated, independent maps for disparity and motion and (iii) a mixture of the two.

First, to characterize how different parameters affected these simulations, we tested a range of columnar arrangements for disparity and motion, different amounts of voxel and neuronal noise, and different relative reliabilities for the disparity and motion cues (**Supplementary Figs. 3** and **4**). We chose parameter values that matched our fMRI data as closely as possible (for example, signal-to-noise ratio) and corresponded to published data (for example, spatial period of disparity representations[17]). These simulations demonstrated the experimental logic, confirming that fused cues surpass quadratic summation (**Supplementary Fig. 3b**) and that independent representations are unaffected by large conflicts and do not support transfer (**Supplementary Fig. 4c**). Second, we explored the composition of the neuronal population, comparing our simulation results to our empirical data (**Fig. 6**). We found a close correspondence between the fMRI decoding data from V3B/KO and a simulated population in which 50–70% of the neuronal population fuses cues (50% for strict fusion, 70% for robust fusion, on the basis of minimizing the χ^2 statistic).

Control analyses

During scanning, we took precautions to reduce the possibility of artifacts. First, we introduced a demanding task at fixation to ensure equivalent attentional allocation across conditions (**Supplementary Fig. 5**). Second, measurements of functional signal-to-noise ratio for each area (**Supplementary Fig. 2c**) showed that differences in prediction accuracy related to stimulus-specific processing rather than the overall fMRI responsiveness. That is, functional signal-to-noise ratio was highest in the early visual areas rather than higher areas that showed fusion.

Finally, eye movements are unlikely to account for our findings. First, although we could not measure eye vergence objectively in the scanner, the attentional task[22] showed that participants maintained vergence well (**Supplementary Fig. 5**) with no reliable differences between conditions. Second, our stimuli were designed to reduce vergence changes: a low spatial frequency pattern surrounded the stimuli, and participants used horizontal and vertical nonius lines to promote correct eye alignment. Together with previous control data using similar disparities[23], this suggests vergence differences could not explain our results. Third, monocular eye movement recordings suggested little systematic difference between conditions (**Supplementary Fig. 6**). Moreover, we found that an SVM could not discriminate near versus far positions reliably on the basis of eye position, suggesting that patterns of eye movement did not contain systematic information about depth positions (**Supplementary Fig. 6**).

DISCUSSION

Estimating three-dimensional structure in a robust and reliable manner is a principal goal of the visual system. A computationally attractive means of achieving this goal is to fuse information provided from two or more signals, so that the composite is more precise than its constituents. Despite considerable interest in this topic, comparatively little is known about the cortical circuits involved. Here we demonstrate that visual area V3B/KO may be important in this process and propose that fusion is an important computation performed by the dorsal visual stream.

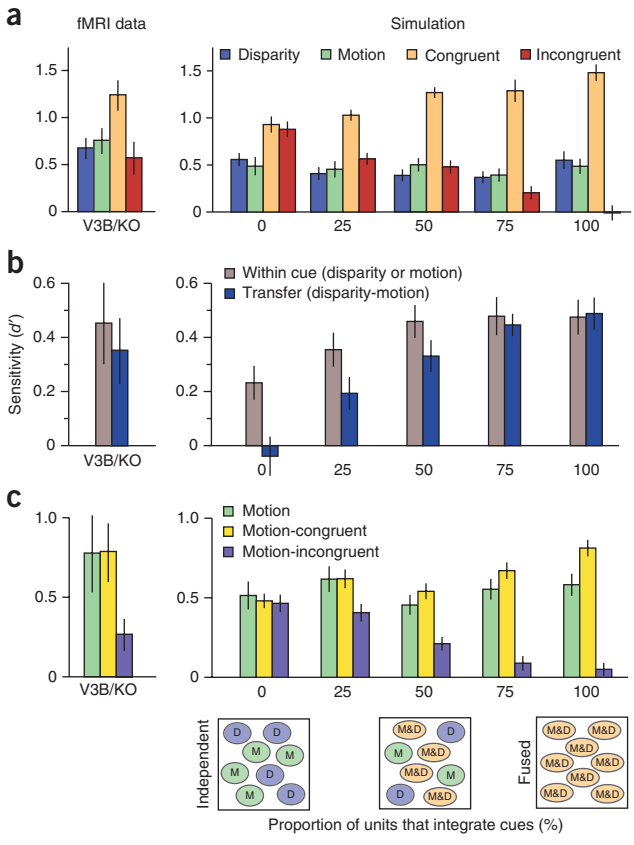

Figure 6 fMRI decoding data from V3B/KO adjacent to results from simulations. (**a**) Simulation results show decoding performance of a simulated population of voxels where the neuronal population contains different percentages of units tuned to individual versus fused cues. The χ^2 statistic was used to identify the closest fit between empirical and simulated data from a range of population mixtures. (**b**) fMRI decoding data for the transfer tests adjacent to the simulation results. (**c**) Performance in a transfer test between data from the motion condition and the consistent and inconsistent cue conditions. Error bars, s.e.m.

First, we showed that fMRI signals from area V3B/KO were more discriminable when two cues concurrently signaled depth, and this improvement exceeded the minimum bound expected for fusion. Second, we showed that improved performance was specific to congruent cues: presenting highly inconsistent disparity and motion information did not improve discriminability. This follows the predictions of integration, and it matched perceptual judgments, but is not expected if disparity and motion signals are colocated but independent. A potential concern is whether the discrimination of brain signals relates to depth *per se*, or to low-level stimulus correlates (for example, speed of movement). We showed that although information about relative motion is diagnostic of depth from disparity, these cross-cue transfer effects are not found between perceptually flat motion and disparity-defined depth. These results suggest a potential neural locus for interactions between disparity and motion depth cues demonstrated in threshold[13] and suprathreshold psychophysical tasks[14,15]. More generally, they highlight V3B/KO as an area that may integrate a range of different signals to estimate depth.

Although our results pointed clearly to area V3B/KO, our quadratic summation, congruent versus incongruent and transfer test analyses all suggested responses in other areas (namely, V3 and V3A) that, although not significant, might also relate to fusion. It is possible that our tests were not sufficiently sensitive to reveal fusion in these (or other) areas for which we have a null result; for instance, decoding accuracies for the motion condition were high in some areas, so responses in the congruent condition may have been near ceiling, limiting detection. However, an alternative is that responses in these earlier areas represent an intermediate depth representation in which links between disparity and motion are not fully established. Previously it was suggested that the kinetic occipital (KO) area is specialized for depth structure[24] and is functionally distinct from V3B. Using independent localizer scans, we do not find a reliable means of delineating V3B from KO. However, to check that we were not mischaracterizing responses, we examined the spatial distribution of voxels chosen by the classifier. We found that chosen voxels were distributed throughout V3B/KO and did not cluster into subregions (**Supplementary Fig. 7**).

Relation between psychophysical and fMRI results

Although results in V3B/KO are consistent with behavioral evidence for fusion, there is a difference in that sensitivity to the 'single' cues differs at the behavioral level (**Fig. 2**) but not at the decoding level (**Fig. 3**). From psychophysical results[13], higher sensitivity to disparity-defined depth is expected. However, this would not necessarily translate to decoding differences. Specifically, our behavioral task measured increment thresholds (sensitivity to small depth differences), whereas fMRI stimuli were purposely[18] suprathreshold (the difference between near and far stimuli was very apparent). Thus, although clear parallels can be drawn between tests for integration at the psychophysical and fMRI levels, necessary differences between paradigms make it difficult to compare the magnitude of the effects directly.

Further, multisensory integration effects for single unit recordings are reported to be highly nonlinear near threshold[25], but more additive or subadditive with suprathreshold stimuli[11,26,27]. Our use of suprathreshold stimuli makes it unsurprising that we did not observe significant changes in overall fMRI responses (**Supplementary Fig. 2**). Moreover, we have not attempted to 'add' and 'subtract' cues (for example, our 'single cue' relative motion stimulus contained disparity information that the viewed display was flat). Our manipulation purposely changes the degree of cue conflict between cues, thereby establishing a minimum bound for fusion. Although useful, testing against this bound alone cannot preclude independence. Specifically, fused cues should have reduced neuronal variability[28]; however, fMRI measures of this activity aggregate responses and are subject to extra noise (for example, participant movement and scanner noise). Depending on the amount of noise, decoding independent representations can surpass the minimum bound (**Supplementary Fig. 3**). The subsequent tests we developed (incongruent cues and transfer test) are therefore important in confirming the results.

Finally, we outlined two variants for the fusion of strongly conflicting cues: strict or robust (**Fig. 1d**). Behaviorally, we found evidence for robust fusion: sensitivity in the incongruent cue condition matched the disparity condition, and perceived depth relied on disparity. This was compatible with fMRI results in V3B/KO, where performance dropped to that seen with 'single' cues. However, we developed a further test of robust fusion: if responses in V3B/KO reflect robust perception, the classifier's predictions might reverse for incongruent stimuli. That is, if depth is decoded at the perceptual level, training the classifier on 'near' motion may predict a 'near' perceptual interpretation of the incongruent stimulus, even though motion signals 'far'. We did not find a reversal of discrimination performance (**Fig. 6c**); however, performance was considerably reduced, suggesting an attenuated response. Although this result *per se* does not match robust fusion, it is compatible with a population mechanism for robust perception. In particular, depth estimation can be understood as causal inference[21] in which the brain computes depth both ways—that is, there is a mixed population that contains both units tuned to independent and to fused cues. A readout mechanism then selects one of the competing interpretations, using the relative reliabilities of the fused and independent models. This idea is compatible with our simulations of a mixed population in V3B/KO, and previous work that suggests V3B/KO is important in selecting among competing depth interpretations[23].

Cortical organization for depth processing

While there is comparatively little work on neural representation of depth from integrated visual cues, individual cues have been studied extensively. Responses to binocular disparity are observed through occipital, temporal and parietal cortices[29,30], and there are links between the perception of depth from disparity and fMRI responses in dorsal and ventral areas[18,31,32]. Similarly, responses to motion-defined depth have been observed in ventral, dorsal and parietal areas[33–35]. To link depth from disparity and motion, previous work has highlighted overlapping fMRI activations[24,36–38]. This suggests widespread cortical loci in which different cues converge; however, this does not imply the shared organizational structure that we demonstrate here.

Our tests of cue fusion reveal V3B/KO as the main cortical locus for depth cue integration. However, tests of motion parallax processing in the macaque have highlighted area MT (also known as V5) (ref. 8). Given well-established disparity selectivity in MT (ref. 17), this suggests a candidate for integrating depth cues. We observed discriminable fMRI responses for both disparity and relative motion in the human MT+ (V5) complex but did not obtain evidence for fusion. While it is possible this represents a species difference[39], the difference may relate to different causes of motion. In particular, we simulated movement of a scene in front of a static observer, whereas previous work[8] moved the participant in a static scene. Thus, in our situation, there was no potential for vestibular signals to contribute to the estimation of ego movement by mediotemporal cortex[10,11].

In interpreting our results, it is important to consider that the MVPA approach we used is generally understood to rely on weak

biases in the responses of individual voxels that reflect a voxel's sample of neuronal selectivities and vasculature[40,41] (although see refs. 42,43). By definition, these signals reflect a population response, so our results cannot be taken to reveal fusion by single neurons. For instance, it is possible that depth is represented in area V3B/KO in parallel for disparity and for motion. However, if this is the case, these representations are not independent: they must share common organizational structure to account for our findings that prediction accuracy falls to single-component levels for incongruent stimuli and that training the classifier on one cue supports decoding of the other. It has been suggested that MVPA decoding of stimulus orientation relies on univariate differences across the visual field[43]. Such spatial organization for disparity preferences has not been identified in the human or macaque brain; however, this is a matter for further investigation. Our previous work[18] and ongoing investigations have not provided evidence of retinotopic disparity organization.

Independence versus fusion

Previously, we tested cue combination by relating psychophysical and fMRI responses[44]. This highlighted the role of ventral lateral occipital cortex in cue combination, which is not the main locus observed here. Differences in stimuli may be responsible: we previously used slanted planes defined by disparity and perspective cues. Thus ventral areas may be more selective for 'pictorial' cues and/or be more selective for slanted surfaces than flat planes. Second, here we used a coarse task, whereas previously[44] we used a fine judgment task that may require greater ventral involvement[30]. However, next we discuss the possibility that the different cortical loci (dorsal versus ventral) point to different types of computation.

In the introduction, we presented two scenarios for optimal judgments: fusion versus independence. Independence increases the separation between classes (for example, 'near' and 'far') but does not reduce variance, whereas fusion reduces the variance of estimates, but leaves separation unchanged. We suggest these two modes of operation may be exploited for different types of task. If a body movement is required, the brain is best served by fusing the available information to obtain an estimate of the scene that is unbiased and has low variance. Such a representation would be particular to the viewing situation (that is, highly specific) and variant under manipulation of individual cues. In contrast, recognition tasks are best served by maximizing the separation of objects in a high-dimensional feature space while ignoring uninformative dimensions. Such a mechanism would support invariant performance by discarding irrelevant 'nuisance' scene parameters, yet may be highly uncertain about the particular structure of the scene[45]. To illustrate the distinction, consider a typical desktop scene. If the observers' goal is to discriminate a telephone from a nearby book, information about the three-dimensional orientation on the tabletop is uninformative, so it should be discounted from the judgment (that is, the telephone's features should be recognized while ignoring location). In contrast, to pick up the telephone, the brain should incorporate all the information relevant to the location from the current view.

Our previous tests of disparity processing[18] suggest differences between the visual pathways: dorsal areas appear selective for metric disparity (that is, the precise location of a plane), whereas ventral lateral occipital cortex represents depth configuration (that is, whether the stimulus is near or far, but not how near or how far). The current findings bolster this suggested distinction by providing evidence for fusion in the dorsal pathway. We propose this provides the best metric information about the scene that is specific to the current view.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

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

ACKNOWLEDGMENTS

We thank B. Tjan, R. Fleming and A. Glennerster for discussions on the project. The work was supported by fellowships to A.E.W. from the Wellcome Trust (095183/Z/10/Z) and Biotechnology and Biological Sciences Research Council (C520620) and to H.B. from the Japan Society for the Promotion of Science (H22,290).

AUTHOR CONTRIBUTIONS

H.B. collected data, programmed stimuli, performed the analysis, wrote the simulations and prepared the work for publication; T.J.P. collected data, programmed stimuli and performed preliminary analysis; A.M. wrote MVPA analysis tools; A.E.W. originated and designed the study, collected data, performed and guided analysis, wrote the simulations, prepared the work for publication and wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Dosher, B.A., Sperling, G. & Wurst, S.A. Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Res.* **26**, 973–990 (1986).
2. Bülthoff, H.H. & Mallot, H.A. Integration of depth modules – stereo and shading. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **5**, 1749–1758 (1988).
3. 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).
4. Clark, J.J. & Yuille, A.L. *Data Fusion for Sensory Information Processing Systems* (Kluwer Academic, 1990).
5. Ernst, M.O. & Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
6. 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).
7. Tsutsui, K., Sakata, H., Naganuma, T. & Taira, M. Neural correlates for perception of 3D surface orientation from texture gradient. *Science* **298**, 409–412 (2002).
8. Nadler, J.W., Angelaki, D.E. & DeAngelis, G.C. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* **452**, 642–645 (2008).
9. 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).
10. Gu, Y., Angelaki, D.E. & DeAngelis, G.C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
11. Morgan, M.L., DeAngelis, G.C. & Angelaki, D.E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–673 (2008).
12. Rogers, B. & Graham, M. Motion parallax as an independent cue for depth perception. *Perception* **8**, 125–134 (1979).
13. Bradshaw, M.F. & Rogers, B.J. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res.* **36**, 3457–3468 (1996).
14. Nawrot, M. & Blake, R. Neural integration of information specifying structure from stereopsis and motion. *Science* **244**, 716–718 (1989).
15. Poom, L. & Borjesson, E. Perceptual depth synthesis in the visual system as revealed by selective adaptation. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 504–517 (1999).
16. Domini, F., Caudek, C. & Tassinari, H. Stereo and motion information are not independently processed by the visual system. *Vision Res.* **46**, 1707–1723 (2006).
17. DeAngelis, G.C. & Newsome, W.T. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* **19**, 1398–1415 (1999).
18. Preston, T.J., Li, S., Kourtzi, Z. & Welchman, A.E. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J. Neurosci.* **28**, 11315–11327 (2008).
19. Nandy, A.S. & Tjan, B.S. Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *J. Vis.* **8**, 3 (2008).
20. 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).
21. Körding, K.P. *et al.* Causal inference in multisensory perception. *PLoS ONE* **2**, e943 (2007).
22. Popple, A.V., Smallman, H.S. & Findlay, J.M. The area of spatial integration for initial horizontal disparity vergence. *Vision Res.* **38**, 319–326 (1998).

23. Preston, T.J., Kourtzi, Z. & Welchman, A.E. Adaptive estimation of three-dimensional structure in the human brain. *J. Neurosci.* **29**, 1688–1698 (2009).
24. Tyler, C.W., Likova, L.T., Kontsevich, L.L. & Wade, A.R. The specificity of cortical region KO to depth structure. *Neuroimage* **30**, 228–238 (2006).
25. Meredith, M.A. & Stein, B.E. Interactions among converging sensory inputs in the superior colliculus. *Science* **221**, 389–391 (1983).
26. Avillac, M., Ben Hamed, S. & Duhamel, J.R. Multisensory integration in the ventral intraparietal area of the macaque monkey. *J. Neurosci.* **27**, 1922–1932 (2007).
27. Stanford, T.R., Quessy, S. & Stein, B.E. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.* **25**, 6499–6508 (2005).
28. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
29. Orban, G.A., Janssen, P. & Vogels, R. Extracting 3D structure from disparity. *Trends Neurosci.* **29**, 466–473 (2006).
30. Parker, A.J. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 379–391 (2007).
31. 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).
32. Chandrasekaran, C., Canon, V., Dahmen, J.C., Kourtzi, Z. & Welchman, A.E. Neural correlates of disparity-defined shape discrimination in the human brain. *J. Neurophysiol.* **97**, 1553–1565 (2007).
33. Orban, G.A., Sunaert, S., Todd, J.T., Van Hecke, P. & Marchal, G. Human cortical regions involved in extracting depth from motion. *Neuron* **24**, 929–940 (1999).
34. 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).
35. Paradis, A.L. *et al.* Visual perception of motion and 3-D structure from motion: an fMRI study. *Cereb. Cortex* **10**, 772–783 (2000).
36. Sereno, M.E., Trinath, T., Augath, M. & Logothetis, N.K. Three-dimensional shape representation in monkey cortex. *Neuron* **33**, 635–652 (2002).
37. Durand, J.B. *et al.* Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron* **55**, 493–505 (2007).
38. Peuskens, H. *et al.* Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J. Cogn. Neurosci.* **16**, 665–682 (2004).
39. 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).
40. Shmuel, A., Chaimow, D., Raddatz, G., Ugurbil, K. & Yacoub, E. Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey information on the stimulated eye. *Neuroimage* **49**, 1957–1964 (2010).
41. Kriegeskorte, N., Cusack, R. & Bandettini, P. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *Neuroimage* **49**, 1965–1976 (2010).
42. Op de Beeck, H.P. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage* **49**, 1943–1948 (2010).
43. Freeman, J., Brouwer, G.J., Heeger, D.J. & Merriam, E.P. Orientation decoding depends on maps, not columns. *J. Neurosci.* **31**, 4792–4804 (2011).
44. Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H. & Kourtzi, Z. 3D shape perception from combined depth cues in human visual cortex. *Nat. Neurosci.* **8**, 820–827 (2005).
45. Tjan, B.S., Lestou, V. & Kourtzi, Z. Uncertainty and invariance in the human visual cortex. *J. Neurophysiol.* **96**, 1556–1568 (2006).

ONLINE METHODS

Observers. Twenty observers from the University of Birmingham participated in the fMRI experiments and thirteen in the psychophysical experiments. Observers had normal or corrected-to-normal vision and were screened for stereo deficits. Experiments were approved by the University of Birmingham STEM ethical review committee; all observers gave written informed consent.

Stimuli. Stimuli were random patterns of black and white dots[18]. A fixation marker was presented at the center of a 1° circular hole in the stimulus and consisted of a square (0.5° on a side) with horizontal and vertical nonius lines (length 0.375°). The random dot region was surrounded by a grid of black and white squares that provided an unambiguous background reference.

We used four different conditions: depth defined by disparity, by motion, by disparity and motion consistently with each other, and by disparity and motion inconsistently with each other. In addition, we created a motion control stimulus. In all cases, a central square (10 × 10°) target plane was presented. The central target was surrounded by a larger rectangle (18 × 14°) of black and white dots (the 'background') for all conditions except in the motion control stimulus, where no background was presented (except for the mid-gray screen). To depict depth from relative motion, the background and target planes moved horizontally following a sinusoidal velocity profile with a period of 1 s (**Fig. 2b**). The background plane movement had amplitude 0.9°, while the target moved with an amplitude of either 1.32° (near) or 0.29° (far). Thus, the relative motion of the target and background gave rise to a pattern of deletion and accretion of the background (near stimuli) or target (far stimuli) dots as the targets and background translated back and forth across the screen. To depict depth from disparity, the central target plane was given a horizontal binocular disparity of ±6 or 9 arcmin while the background was presented in the plane of the screen. For the disparity-defined depth stimulus, the whole stimulus (target and background) moved rigidly with a sinusoidal horizontal movement (amplitude 0.9°, period 1 s). In contrast, for stimuli depicting disparity and motion-defined depth, the central target plane had ±6 or 9 arcmin disparity and a movement amplitude of either 1.32° or 0.29°. Differences in motion amplitudes for motion-defined depth produced a difference in mean speed for near and far depth positions. To assess the impact of this speed difference, the speed control stimulus contained only the central square (no random dot background) moving with an amplitude of either 1.32° or 0.29°. Without movement of the background, this stimulus yielded no impression of depth. Stereoscopic presentation and display parameters matched previous work[18]. To control for attention and promote proper fixation, observers performed a subjective assessment of eye vergence[22]. Vernier targets were flashed for 250 ms at either side of the desired fixation position, and a logistic function was fit to the proportion of "target on the right" responses as a function of the vernier displacement.

Psychophysics. Behavioral tests were conducted in the lab using a stereoscope in which the two eyes viewed separate CRTs (ViewSonic FB2100x) through front-silvered mirrors. Stimulus parameters were equivalent (in terms of visual angle) to those used for scanning. Participants judged which of two, sequentially presented stimuli had the greater depth (presentation time, 1 s; interstimulus interval, 1 s). A standard stimulus (depth specified by ±6 arcmin and/or movement amplitude of 1.32° or 0.29°) was presented on every trial. The other stimulus contained a depth increment (disparity and/or motion) relative to the standard, whose magnitude was controlled by a staircase algorithm. Participants judged "which was closer" for near targets and "which was farther" for far targets. Four conditions were randomly interleaved during each experimental run: motion, disparity, congruent cues, incongruent cues. In the incongruent case, disparity specified near and motion far, or vice versa. The increment applied to the stimulus was therefore an increase in depth away from the fixation plane, but in opposite directions for the two cues (for example, Δdisparity was nearer, and Δmotion was farther). Just noticeable difference thresholds were estimated from 60 staircase trials per condition. There were 480 trials per run (60 trials × 4 conditions × 2 depth positions: near or far). Each participant's thresholds were measured 2–4 times and then averaged. In a separate experiment, we measured the perceptual interpretation of the stimuli by presenting near or far stimuli from each condition 20 times and asking participants to indicate whether the stimulus was near or far. The informative result was that participants relied on disparity when judging the inconsistent-cue stimulus (100% reliance on disparity for every participant).

Imaging. Data were acquired at the Birmingham University Imaging Centre using a 3-tesla Philips MRI scanner with an eight-channel head coil. Blood oxygen level–dependent signals were measured with an echo-planar sequence (TE, 35 ms; TR, 2,000 ms; 1.5 × 1.5 × 2 mm, 27 or 28 near coronal slices) for both experimental and localizer scans. A high-resolution anatomical scan (1 mm) was also acquired for each participant. Four separate experiments were run (eight, seven, four and five participants, respectively); each had four stimulus types (a subset of the following conditions: disparity, relative motion, disparity and motion consistent, disparity and motion inconsistent or motion control) in two configurations (near and far) and a fixation baseline condition.

Stimuli were presented in blocks of 16 s (blocked fMRI design). In each block, stimuli were picked randomly from a set of 24 example stimuli (per subject) that differed in the random placement of dots making up the stereogram. Individual stimuli were presented for 1 s, followed by a 1-s fixation period. Three blocks of each stimulus type were presented during an individual run in a counterbalanced randomized order, and the scan started and ended with a 16-s fixation interval. Scans lasted a total of 416 s. Eight runs were collected for each observer.

For each observer, regions of interest (ROIs) in visual cortex were defined using standard retinotopic mapping procedures (described elsewhere[18]). Area V3B/KO (ref. 46) was defined as the set of contiguous voxels located anterior to V3A, inferior to V7 and posterior to the human motion complex (human MT+/V5) that responded significantly more highly ($P < 10^{-4}$) to kinetic boundaries than to transparent motion of a field of black and white dots.

We used BrainVoyager QX (BrainInnovation B.V.) to transform anatomical scans into Talairach space, inflate the cortex and create flattened surfaces of both hemispheres for each subject. Each functional run was preprocessed using three-dimensional motion correction, slice time correction, linear trend removal and high-pass filtering (three cycles per run cut-off). No spatial smoothing was performed on the functional data used for the multivariate analysis. Functional runs were aligned to the subject's corresponding anatomical scan and transformed into Talairach space.

Multi-voxel pattern analysis. Within each ROI, we sorted gray matter voxels according to their response (t-statistic) to all stimulus conditions in comparison to fixation baseline across all experimental runs. This procedure resulted in the selection of 250 voxels per ROI[18]. We normalized (z-score) the time course of each voxel separately for each experimental run to minimize baseline differences between runs. Test patterns for the multivoxel analysis were generated by shifting the fMRI time series by 4 s to account for the hemodynamic response lag. To control for the possibility that classification accuracy was due to a univariate response to a particular volume, we normalized the mean of each data vector for each volume to zero by subtracting the mean over all voxels for that volume[47]. In this way the data vectors for each volume had the same mean value across voxels and differed only in the pattern of activity. We used a linear support vector machine (SVM^{light} toolbox) for classification and performed an eightfold leave-one-out cross validation in which data from seven scans were used as training patterns (21 patterns, 3 per run) and data from the remaining run was used as test patterns (3 patterns). For each subject, we took the mean accuracy across cross-validations. We converted prediction accuracies into units of discriminability (d') using the formula

$$d' = 2 \times \mathrm{erfinv}(2p - 1) \quad (4)$$

where erfinv is the inverse error function and p the proportion of correct predictions.

To conduct the transfer test analysis, we used a recursive feature elimination method[48] to detect sparse discriminative patterns and define the number of voxels for the SVM classification analysis. In each feature elimination step, a small proportion of voxels was discarded until there remained a core set of voxels with the highest discriminative power. To avoid circular analysis, the recursive feature elimination method was applied independently to the training patterns of each cross-validation, resulting in eight sets of voxels. This was done separately for each experimental condition, with final voxels for the SVM analysis chosen on the basis of the intersection of voxels from corresponding cross-validation folds. A standard SVM was then used to compute within- and between-cue prediction accuracies. This feature selection method was required to provide robust evidence of transfer, in line with previous evidence that it improves generalization[48].

doi:10.1038/nn.3046

Statistical analysis was performed in SPSS (IBM Corporation), and Greenhouse-Geisser correction was used when appropriate.

Simulations. Using Matlab (Mathsworks Inc.), we simulated a population of 'depth columns', each of which had a mean depth preference and Gaussian tuning profile. These maps had a sawtooth structure whose phase progression was randomly perturbed to create jittered maps[49]. The cycle width of depth representation was set at 3 mm (scaling by a factor of 2 from the macaque[17]), although we tested for generality with other scales (**Supplementary Fig. 4**). We considered two main population models: fusion and independence (separate maps for disparity and motion). Under independence, maps for disparity and motion were assumed correlated, but with some jitter, and were sampled irregularly by each voxel (see **Supplementary Fig. 4** for an investigation of these parameters). We also considered mixed populations by varying the proportion of columns responding to fused versus single cues (**Fig. 6**). Column tuning width was set at $\sigma_d = \sigma_m = 12$ arcmin for single cues (σ_d, s.d. for disparity; σ_m, s.d. for motion) and the integrated response followed maximum likelihood estimation (s.d., $\sigma_i = 8.49$), although we tested generality using other values (**Supplementary Fig. 3c**). We convolved the stimulus (Gaussian, $\sigma = 0.2$ arcmin) and the column tuning profile to calculate the pattern of neuronal activity evoked by the stimulus. This response was subject to a compressive nonlinearity and added noise (**Supplementary Fig. 3**). To calculate voxel responses, we averaged the responses of individual columns that were sampled by a coarser scale voxel grid. These aggregated column responses were then subjected to 'voxel noise' (**Supplementary Fig. 3**). We investigated the contribution of different signal-to-noise ratios for neural (0.4 to 4.5) and fMRI (0.6 to 1.5) responses (**Supplementary Fig. 3**) and chose a value for the functional signal-noise-ratio that matched the empirical data from V3B/KO (0.93). We used the same SVM analysis tools to decode the simulated data as were used for the empirical data. We simulated 250 voxels with 8 runs of 24 patterns for both near and far presentations for each condition (that is, the dimensionality of the empirical study). The SVM classifications were repeated for each of the 20 participants in the fMRI experiment, and we then calculated the between-subjects average and s.e.m.

46. Dupont, P. *et al.* The kinetic occipital region in human visual cortex. *Cereb. Cortex* **7**, 283–292 (1997).
47. Serences, J.T. & Boynton, G.M. The representation of behavioral choice for motion in human visual cortex. *J. Neurosci.* **27**, 12893–12899 (2007).
48. De Martino, F. *et al.* Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* **43**, 44–58 (2008).
49. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8**, 679–685 (2005).