Fusion of visual cues is not mandatory in children

Marko Nardini[a,b,1], Rachael Bedford[b,2], and Denis Mareschal[b]

[a]Department of Visual Neuroscience, University College London Institute of Ophthalmology, London EC1V 9EL, United Kingdom; and [b]Centre for Brain and Cognitive Development, Birkbeck, University of London, London WC1E 7HX, United Kingdom

Edited by Brian A. Wandell, Stanford University, Stanford, CA, and approved July 27, 2010 (received for review February 16, 2010)

Human adults can go beyond the limits of individual sensory systems' resolutions by integrating multiple estimates (e.g., vision and touch) to reduce uncertainty. Little is known about how this ability develops. Although some multisensory abilities are present from early infancy, it is not until age ≥8 y that children use multiple modalities to reduce sensory uncertainty. Here we show that uncertainty reduction by sensory integration does not emerge until 12 y even within the single modality of vision, in judgments of surface slant based on stereoscopic and texture information. However, adults' integration of sensory information comes at a cost of losing access to the individual estimates that feed into the integrated percept ("sensory fusion"). By contrast, 6-y-olds do not experience fusion, but are able to keep stereo and texture information separate. This ability enables them to outperform adults when discriminating stimuli in which these information sources conflict. Further, unlike adults, 6-y-olds show speed gains consistent with following the fastest-available single cue. Therefore, whereas the mature visual system is optimized for reducing sensory uncertainty, the developing visual system may be optimized for speed and for detecting sensory conflicts. Such conflicts could provide the error signals needed to learn the relationships between sensory information sources and to recalibrate them while the body is growing.

sensory | development | integration | vision | depth perception

Human adults can reduce sensory uncertainty by integrating estimates, both across modalities (1, 2) (e.g., integrating vision and touch to judge size) and within a modality (3) (e.g., integrating visual stereoscopic and texture information to judge surface slant). Given independent estimates with uncorrelated Gaussian noise, the optimal reduction in uncertainty (variance) is obtained by a weighted averaging of estimates, in which each estimate is weighted in proportion to its relative reliability (1/variance) (4, 5). Whereas human adults can achieve the optimal level of variance reduction in sensory tasks (1, 2), the developmental time course of this ability is unclear. Although some multisensory abilities are present from early infancy (6–8), in recent studies, children did not integrate information across modalities for shape discrimination, spatial localization, or detection of visual–auditory events until age ≥8 y (9–11). In adults, sensory integration can lead to mandatory "fusion," in which the ability to judge the individual component estimates is lost (12, 13). Sensory fusion is especially strong for information within a single modality (12). One hypothesis for late development of integration is that keeping information separate is adaptive in allowing senses to be calibrated against each other while the body is growing (10, 14). To test whether children do keep sensory information sources separate, we tracked the development of sensory integration and fusion within the single modality of vision.

Results

Experiment 1: Cue Integration. The gradient of change in element size and density in a homogenously tiled surface's projection ("texture") provides information about its slant (15). A second source of slant information comes from binocular disparity (16). Adults integrate texture and disparity cues to reduce the uncertainty of their estimates of slant, adjusting their relative weightings as they change in reliability (3). We studied the development of uncertainty reduction and reweighting, both markers of mature sensory integration (4, 5), for these two visual cues to surface slant. To rapidly measure young children's thresholds and weightings we used staircase procedures with <200 total trials per participant. First, we measured sensitivity to slant differences signaled by disparity alone (condition *D*), texture alone (condition *T*), and both together (condition *DT*) (Fig. 1*A*). In each condition, children and adults judged which of two planes appeared "flattest" to the ground. Mean 75% discrimination thresholds declined with age, showing improving sensitivity to slant in all conditions [Fig. 1*B*; linear effects of *age group* in ANOVA, $F(1, 4) = 3.4$, $P < 0.001$ for *D*; $F(1, 4) = 2.3$, $P < 0.01$ for *T*; $F(1, 4) = 5.8$, $P < 0.001$ for *DT*]. If estimates are integrated to reduce uncertainty, *DT* will show a lower threshold than either single-cue condition. Twelve-year-olds' and adults' thresholds were significantly lower for *DT* than for either single cue [Fig. 1*B*, **; 12 y, $t(18) = 2.1$, $P = 0.047$ vs. *D*; $t(18) = 2.2$, $P = 0.039$ vs. *T*; adults, $t(18) = 2.5$, $P = 0.020$ vs. *D*; $t(18) = 5.8$, $P < 0.001$ vs. *T*]. These groups' *DT* thresholds also matched optimal (ideal observer) predictions (Fig. 1*B*, *SI Materials and Methods*, and Eq. **S3**).

Therefore, observers aged ≥12 y integrated disparity and texture information optimally to reduce their uncertainty in judging surface slant. Younger participants did not have significantly lower thresholds given *DT* vs. their best single cue, although all groups (especially 8-y-olds) showed a trend for improvement relative to both cues (Fig. 1*B*). This trend is consistent with a minor advantage for binocular viewing of monocular (texture gradient) information, but could also potentially indicate integration masked by noisy responding. Results from experiment 3, in which we were able to quantify children's "lapse rate" with the same task and stimuli, indicate that lapse rate was extremely low, making the latter interpretation highly unlikely (*Results, Experiment 3: Latency*). Variability across observers was also sufficiently low to show significant differences between measured *DT* thresholds and ideal observer predictions at 6 and 10 y [at 6 y, $t(18) = 3.1$, $P = 0.006$; at 10 y, $t(17) = 2.8$, $P = 0.012$] (Fig. 1*B*).

To assess integration further we examined a second marker for it, reweighting by reliability, in the same observers. Because optimal variance reduction entails weighting estimates with respect to their reliabilities (4, 5), optimal observers must reweight estimates when their reliabilities change. The differences in a plane's projection caused by a fixed (e.g., 5°) change to its slant are greatest for planes with high slant (15, 17). Consequently, as a plane's slant increases, observers can detect increasingly fine changes in this slant on the basis of texture (refs. 3 and 18 and Fig. S1). Base slant has much less effect on disparity-based discrimination (3). Observers taking account of reliability should therefore weight texture more for near-horizontal ("high slant") than near-vertical ("low slant") planes. To assess observers' weighting for (reliance on) disparity

Author contributions: M.N., R.B., and D.M. designed research; R.B. performed research; M.N. and R.B. analyzed data; and M.N. and D.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: m.nardini@ucl.ac.uk.

[2]Present address: Centre for Research in Autism and Education, Institute of Education, London WC1H 0AA, United Kingdom.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001699107/-/DCSupplemental.

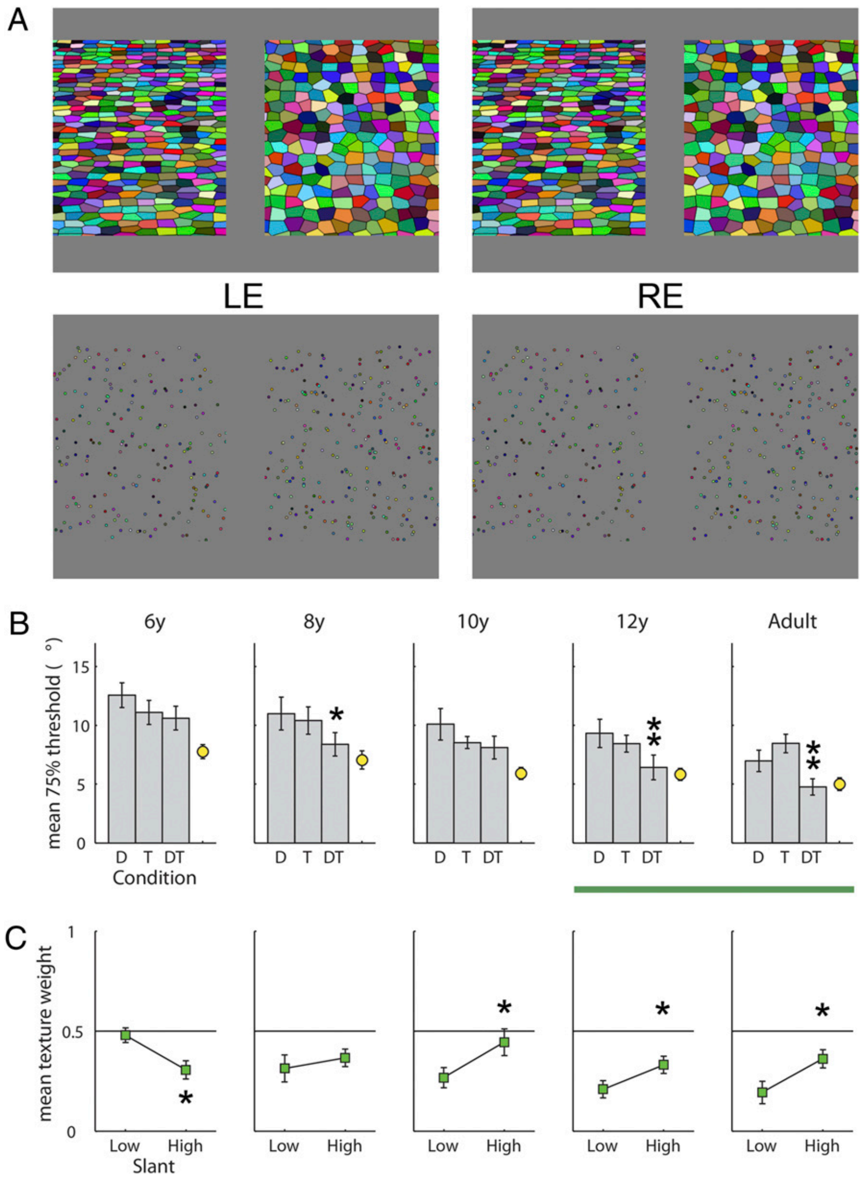

Fig. 1. Cue integration experiment. (*A*) Left eye (LE) and right eye (RE) views of planes with disparity and texture (*Upper*, condition *DT*) and disparity-only (*Lower*, *D*) cues to slant. To view stereoscopically, free fuse by diverging the eyes (slants from disparity may differ from those in the experiment, where these images took up 13° of visual angle). For texture-only (*T*) trials, tiled planes were viewed monocularly. (*B*) Mean ± SEM 75% discrimination thresholds by age group (total $n = 92$). Circles: mean ± SEM ideal observer predictions for *DT*. *DT* thresholds significantly lower on two-tailed paired *t* test than either *D* or *T* are shown by * and significantly lower than both *D* and *T* by **. (*C*) Mean ± SEM relative weighting for texture vs. disparity in conflict conditions with low and high slant. Differences significant at the 5% level are shown by *. Green lines: groups showing mature behavior on each measure.

vs. texture, we measured the point at which they judged a plane with congruent disparity and texture to have the same slant as a plane in which disparity and texture conflict by 10°. Ten- and 12-y-olds and adults weighted texture significantly more for high base slants (Fig. 2*C*), consistent with taking reliability into account [10 y, $t(17) = 2.9, P < 0.01$; 12 y, $t(18) = 2.5, P = 0.02$; adults, $t(18) = 3.1, P < 0.01$]. By contrast, 8-y-olds showed no significant difference in weighting [$t(16) = 0.7, P = 0.51$], whereas 6-y-olds' weighting was in the opposite direction, relying on texture more when it is less reliable [$t(18) = -3.0, P < 0.01$]. These younger groups' behavior indicates that they either misperceived the cues' reliabilities or used reliability information in a manner different from that required to weight cues for optimal uncertainty reduction.

In sum, both markers of mature integration—lower discrimination thresholds given two cues vs. either one (Fig. 1*B*) and reweighting by reliability (Fig. 1*C*)—were first present at 12 y. Six-year-olds showed neither ability, whereas 8- and 10-y-olds were transitional in showing, respectively, a (nonsignificant) trend toward a lower threshold without reweighting, and reweighting without a lower threshold. Sensory integration for uncertainty reduction therefore did not emerge any earlier within the single modality of vision than across modalities in previous studies (9, 10).

Experiment 2: Fusion. In adults, integrating sensory estimates can lead to "sensory fusion" in which observers lose access to the individual estimates that feed into the integrated percept (12, 13). As a consequence, some stimuli that can be discriminated using single cues become metameric (perceptually indistinguishable) when a second cue is added. However, if children do not integrate visual information, they should not be subject to fusion. To test this

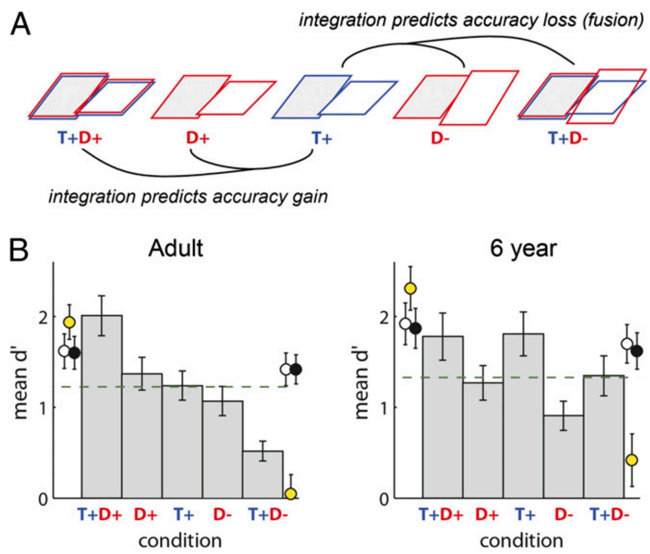

Fig. 2. Fusion experiment. (*A*) Schematic view of five experimental conditions, in which observers made same/different slant judgments for pairs of planes, a standard (shown gray) and a comparison (shown white). On "different" trials (shown), the comparison plane's slant was greater (+) or less (−) than the standard plane's slant, based on single or combined texture (*T*) and disparity (*D*) cues. On "same" trials, standard and comparison planes had the same slants, signaled by these same cues. Integration predicts improved discrimination given congruent combined cues vs. the best single cue, but reduced discrimination ("fusion") given incongruent combined cues vs. the worst single cue. (*B*) Mean $d' \pm$ SEM, measuring discrimination for same vs. different slants by condition in adults (n = 20) and 6-y-olds (n = 20). Combined-cue conditions *T+D+* and *T+D−* are compared with mean ± SEM *d′* predicted by integration (yellow circles) and probability summation of single cues (white and black circles). The probability summation model using a "liberal" criterion detects a difference if *either* single cue signals a difference (white circles). The model using a "conservative" criterion detects a difference only if *both* single cues signal a difference (black circles). See *SI Materials and Methods* for details.

prediction, we compared 6-y-olds' and adults' abilities to judge whether the slants of two elliptical discs (Fig. S2) were the same or different, on the basis of single or combined cues. In single-cue conditions $D+$, $T+$, and $D-$, differences in a comparison plane's slant relative to a 45° standard plane were signaled by disparity (D) or texture (T) cues to greater (+) or lesser (−) slant (Fig. 2*A*). In combined-cue conditions, differences in the two planes' slants were signaled by combinations of these single cues, which were either congruent (condition T+D+) or conflicting (condition T+D−) (Fig. 2*A*). These differences were ±12.5° for adults and ±25° for 6-y-olds. At these levels, mean discrimination (d') (19) for same vs. different slants over single-cue conditions $D+$, $T+$, and $D-$ was matched across groups (Fig. 2*B*, dashed lines). We compared performance on two-cue conditions D+T+ and D+T− with model predictions for integration of cues (Fig. 2*B*, yellow circles) and for use of single cues by probability summation (Fig. 2*B*, white and black circles; Fig. 2*B* legend, *SI Materials and Methods*, Fig. S3, and Eqs. **S4–S7**).

To assess any improvements in discrimination ability given integration of congruent cues, we compared congruent two-cue condition T+D+ with its component single-cue conditions T+ and D+. Adults' mean d' was significantly higher for T+D+ than for either single cue [Fig. 2*B*; $t(19) = 3.3$, $P < 0.01$ vs. D+; $t(19) = 3.6$, $P < 0.01$ vs. T+] and matched model predictions for integration of cues (Fig. 2*B*, yellow circle). Six-year-olds' mean d' was similar to that for the single best cue, did not match predictions for integration (Fig. 2*B*, yellow circle), but was consistent with a probability summation model predicting responses based on single cues [Fig. 2*B*, black circle; $t(19) = 0.6$, $P = 0.59$]. Thus, adults but not 6-y-olds showed improved slant discrimination consistent with integrating congruent cues.

To assess any reduction in discrimination ability given integration of conflicting cues (fusion), we compared conflicting condition T+D− with its component conditions T+ and D−. Adults' mean d' was significantly lower for T+D− than for either single cue [Fig. 2*B*; $t(19) = 4.5$, $P < 0.001$ vs. T+; $t(19) = 3.6$, $P < 0.01$ vs. D−], showing that they were subject to fusion. Stimuli in this condition were marginally more discriminable than complete fusion would predict [Fig. 2*B*, yellow circle; $t(19) = 2.0$, $P = 0.056$], suggesting that adult observers retained some access to single cues (as in ref. 12). However, their discrimination was much poorer than predicted by responding using the individual component cues (Fig. 2*B*, black and white circles). Six-year-olds' mean d' in this same condition (T+D−) was not lower than the worst single cue, but intermediate to the two, indicating that they were not subject to fusion. Their performance was consistent with a probability summation model predicting responses based on the single-component cues [Fig. 2*B*, black circle; $t(19) = 1.0$, $P = 0.32$], but not consistent with fusion [Fig. 2*B*, yellow circle; $t(19) = 3.3$, $P < 0.01$].

In sum, adults' performance indicated integration of cues, whereas 6-y-olds' performance was consistent with responding on the basis of single cues. Whereas integrating cues conferred an advantage on adults when the cues agreed (congruent condition T+D+), not integrating cues conferred an advantage on 6-y-olds when the cues disagreed (conflicting condition T+D−). Thus, by not integrating disparity and texture information, 6-y-olds were able to make some kinds of perceptual judgments that adults found near impossible.

Experiment 3: Latency. We next considered whether 6-y-olds' ability to keep cues separate may also be accompanied by differences in the time course of their responding to single vs. combined cues. Whether given single or multiple information sources, an observer must decide when to stop collecting information and respond. Such a "decision rule" can be optimized for accuracy, speed, reward, or other goals (20). Whereas adults' use of multiple information sources is often optimal for accuracy, current and previous (9, 10) results indicate that children's is not. In experiment 3 we asked whether children use a decision rule that is optimized for speed. An observer given two information sources rather than one can make speed gains by preferentially responding on the basis of the first available source on each trial, as the average time of winners of races is shorter than the average time of any single contestant (21). Thus, if children keep disparity and texture information separate, they may show speed gains given both at the same time vs. either one alone. By contrast, if adults are integrating the estimates, they may have to wait for the slowest single estimate.

We tested discrimination for planes (Fig. 1*A*) at several constant levels of slant difference. Mean overall proportions of correct responses were similar for adults (mean = 0.80, SEM = 0.01) and 6-y-olds [mean = 0.76, SEM = 0.02; $t(30) = 1.8$, $P = 0.09$], showing that the tasks were comparable in difficulty. At intermediate levels not near floor or ceiling, adults were significantly more accurate given both disparity and texture together (DT) than either D or T alone [Fig. 3; at 5°, $t(16) = 5.4$, $P < 0.001$ vs. D, $t(16) = 3.6$, $P < 0.01$ vs. T; at 7°, $t(16) = 2.7$, $P = 0.01$ vs. D, $t(16) = 2.3$, $P = 0.04$ vs. T]. Children's DT accuracy was intermediate to D and T at all levels (Fig. 3). Thus, adults improved in accuracy given two information sources, consistent with integration, whereas children showed no improvement, consistent with responding on the basis of single cues.

The latencies with which observers made these decisions (Fig. 3) showed a different pattern. Mean overall response latencies did not differ for adults (mean = 1.54, SEM = 0.08 s) and 6-y-olds [mean = 1.62, SEM = 0.07; $t(30) = 0.8$, $P = 0.46$]. How-

PSYCHOLOGICAL AND COGNITIVE SCIENCES

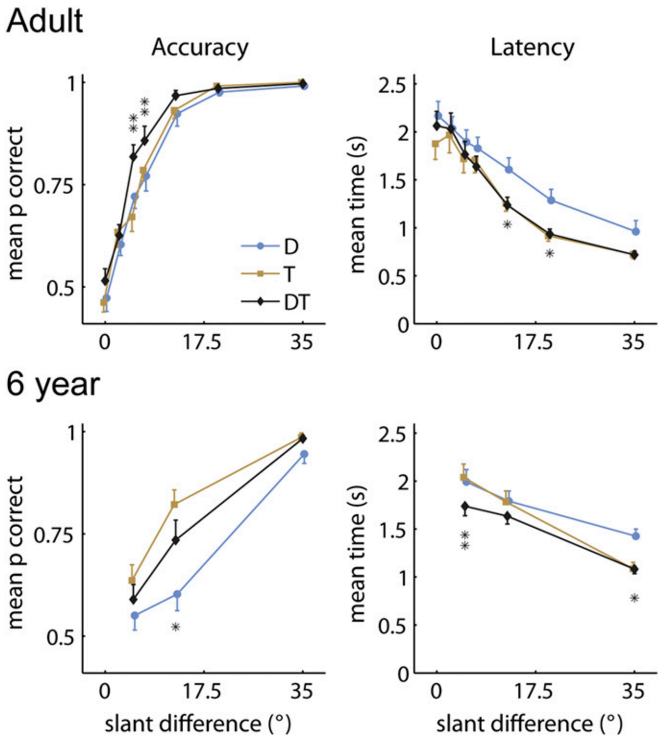

Fig. 3. Latency experiment: Mean ± SEM proportions of correct judgments of the greater slant (*Left*) and mean ± SEM response latencies (*Right*) for adults (n = 17) and 6-y-olds (n = 15), comparing pairs of planes defined by disparity (*D*), texture (*T*), or both (*DT*). *DT* mean differs significantly from either *D* or *T* (*) or from both *D* and *T* (**) on two-tailed paired t test.

ever, whereas adults were never faster given two cues (*DT*) than either one alone, 6-y-olds were significantly faster given two cues than either single cue at their most difficult stimulus level, 5°; $t(14) = 3.0, P = 0.02$ vs. D, $t(14) = 2.5, P = 0.03$ vs. T. Therefore, 6-y-olds exploited multiple cues for speed, tending to follow the fastest-available single cue on each trial. By contrast, if the component disparity and texture estimates are not individually accessible to adults in stimuli including both information sources (experiment 2 and ref. 12), their responses cannot be based on a race (21) between single estimates. Adults' responses may be the outcome of an evidence accumulation process (22, 23) integrating the two. Such a process could reach a decision bound faster given two information sources than one. This may explain why adults were (nonsignificantly) faster with *DT* than with the slowest single cue on those levels at which they made accuracy gains consistent with integration. At easier levels, both groups showed significant speed gains for *DT* relative to *D*, but not *T* (Fig. 3), consistent with simply relying on texture. Texture was much faster than disparity at the easier stimulus levels (Fig. 3), in line with texture-based slant discriminations becoming progressively easier toward the horizontal (15) and with binocular fusion for stereopsis requiring additional time irrespective of stimulus difficulty. An explicit model of speed and accuracy for decisions based on integrated vs. single cues needs to be developed to test this interpretation further.

Experiment 3 also enabled us to estimate "lapse rate" (the rate of error owing to "noisy" or inconsistent responding). This estimation allows us to assess whether such responding could have contributed to results in experiment 1, which used an identical task with the same stimuli. With the easiest stimulus, texture at 35°, 6-y-olds were correct on >98% of trials. The lapse rate must therefore be <2%. Thus it is highly unlikely that noisy responding (e.g., arising from inattention or poor understanding of the task) could have contributed significantly to children's results in experiment 1.

Discussion

A major finding across all three experiments is that whereas adults gained sensitivity to slant given disparity and texture together vs. either one alone, 6-y-olds did not. This result was found using a staircase procedure (experiment 1), same/different judgments (experiment 2), and the method of constant stimuli (experiment 3). Children's failures to show accuracy gains given multiple cues are not likely to be an artifact of inconsistent or noisy responding, because 6-y-olds were capable of responding with >98% accuracy given easy stimuli in experiment 3. Mature sensory integration based on weighting by reliability was not seen until 12 y. Therefore, the ability to reduce uncertainty by integrating information within a single modality does not necessarily develop any earlier than between modalities (9, 10). Recent models propose that uncertainty reduction depends on neural computations implementing weighted averaging of sensory estimates (24, 25). The networks relevant for the present task are those linking early retinotopic visual areas with dorsal and ventral extrastriate areas hMT+/V5 and LOC, associated with combined texture- and disparity-based representations of slant (26). Immaturities on our task could result from pathways to these higher visual areas still developing and processing information in ways not characteristic of the adult visual system (27). Normal early sensory experience appears crucial for development of cross-modal perception (28, 29) and is likely also to be important for development of cue integration within vision. The developmental trajectories for cue weighting and cue integration between 6 and 12 y may be uneven (Fig. 1) and may be investigated further using both psychophysics and neuroimaging.

We found two further striking results. First, 6-y-olds were not confused by conflicting stimuli with which adults experience sensory fusion, but remained able to make judgments on the basis of the component cues. Fusion entails a poor ability to detect sensory conflicts (13). Such conflicts provide an error signal that could be used to calibrate sensory information sources against each other. Being able to detect these information sources separately could be useful for initial learning of the correct mapping between disparity and texture slant information. Remaining able to detect sensory conflicts into late childhood could also be adaptive for growing children. For example, changing interocular distance alters the relationship between disparity and texture cues to slant. These are two reasons why not integrating cues might be adaptive for children, although the underlying hypothesis that integrating cues precludes the ability to recalibrate them (10, 14) still needs to be tested directly.

Second, unlike adults, 6-y-olds showed speed gains given difficult judgments based on two cues. Thus, whereas adults are optimized for accuracy, children may be optimized for speed. Importantly, despite their speed gains, children's mean latencies were no faster than adults'. Therefore, by prioritizing speed, children merely kept up with adult speeds. The coupling of vision and action must be fast enough to keep up with events in the world (30). Given their slower information processing (31), it would be adaptive for children to prioritize speed in general if this strategy enables them to act quickly enough to keep up with real-world events.

These results suggest that developing and mature visual systems are optimized for different goals. Whereas the mature visual system is optimized for reducing sensory uncertainty, the developing visual system may be optimized for speed and for detecting sensory conflicts. Such conflicts may provide the error signals needed to learn the relationships between sensory information sources and to recalibrate them while the body is growing.

Materials and Methods

Stimuli were presented on a 22-in color CRT monitor (Mitsubishi Diamond Pro-2070SB) at 1,152 × 864 pixels, using Psychophysics Toolbox for Matlab (32). Participants viewed stimuli from 175 cm using a chin rest, wearing shutter glasses presenting separate images to the two eyes (CrystalEyes 3; StereoGraphics), each refreshed at 60 Hz. The width of the whole display (Fig. 1*A*) was 13° of visual angle. For monocular conditions participants also wore a patch over one eye. Stereoscopic stimuli were projected with respect to each observer's interocular distance. Participants, who were recruited from a database of volunteers and from schools in and near London, were tested with parents' or their own informed consent in line with ethics committee guidelines. Participants were not given feedback on their performance, but children were rewarded with stickers during breaks to maintain their interest in the task. Participants who failed the TNO stereo test (33), performed at chance, or had thresholds above the measurable range were excluded.

Experiment 1: Cue Integration. Participants were aged 6, mean (SD) age = 6.5 (0.3) y, n = 19 (9 male); 8, mean (SD) age = 8.5 (0.3) y, n = 17 (9 male); 10, mean (SD) age =10.4 (0.3) y, n = 18 (10 male); 12, mean (SD) age = 12.3 (0.2) y, n = 19 (10 male); or were adult, mean (SD) age = 25.8 (4.4) y, n = 19 (7 male). All participants passed the TNO test for stereo vision (33) or were correct more often than chance (binomial test) on disparity-only (*D*) trials. Observers viewed planes (Fig. 1*A*) made of colored tiles binocularly (condition *DT*) or monocularly (condition *T*) and planes made of uniform dots with only disparity cues to slant binocularly (condition *D*). To judge which plane is closest to the horizontal, observers were asked to imagine that they are looking at two roads and to indicate which would be easiest to walk along. The stimuli are described in detail in *SI Materials and Methods*.

In conditions *D*, *T*, and *DT* we used a staircase procedure to measure observers' 75% thresholds for detecting the more slanted plane, given a 45° standard and a simultaneously presented comparison slanted >45°. Each condition began with practice trials at level 25° (i.e., judging a difference between 45° and 70° slants, Fig. 1*A*), which ran until four were correct. Next, slant differences decreased in steps of 4° until either an incorrect response was made or the difference descended to 1°. The final level from this phase was the starting point for a weighted up/down staircase (34) with a 1:3 ratio (1° vs. 3°) between "up" and "down" steps. This staircase converges on the 75% threshold (34). Each staircase had 30 trials, with a break after 15. The upper and lower limits of the staircase were 25° and 0°. To estimate a 75% threshold for each observer and condition we used least-squares fits of psychometric functions to the collected responses (*SI Materials and Methods*). Thresholds near or above the staircase's upper limit, 25°, could not be measured. We therefore excluded participants with an estimated threshold >23.5° in any condition from analysis (n = 9; six aged 6 y, four high disparity-only and two high texture-only thresholds; two aged 8 y, both high texture-only thresholds; and one aged 12 y, high disparity-only threshold).

To assess cue weighting on low slant and high slant conditions, we measured the point of subjective equality (PSE) for a standard plane in which slant from disparity and texture disagreed and a comparison in which these cues agreed (35). Standard planes had slant 30° from texture, 20° from disparity (low slant) or 70° from texture, 60° from disparity (high slant). On the first trial, the comparison plane's slant was the average of the standard plane's two conflicting cues (25° for low slant, 65° for high slant). On each trial observers judged which plane had greatest slant. We used a one-up-one-down staircase of 30 trials with 0.5° steps to make the comparison plane less slanted whenever observers judged it more slanted and vice versa. This staircase converges on a level at which the consistent plane is on average judged the same in slant as the conflicting plane. To estimate this point, which corresponds to observers' relative weightings for the two cues, we used a least-squares fit to the collected responses (*SI Materials and Methods*). There were thus five experimental conditions in total (*D*, *T*, *DT*, low slant, and high slant), whose order for each participant was random.

The ideal observer prediction for condition *DT* (Fig. 1*B*, circles) was formulated assuming that (*i*) texture and disparity noise is uncorrelated and (*ii*) texture and disparity information available in single-cue conditions *D* and *T* provides as good a basis for judging slant as the information available in two-cue condition *DT*. Our assumptions and approach are based on previous more detailed validation of similar stimuli (3). The ideal observer prediction and the prediction that texture information will be most reliable in planes with high slant are described in detail in *SI Materials and Methods*.

Experiment 2: Fusion. Participants, all showing stereo vision on the TNO test (33), were aged 6, mean (SD) age = 6.5 (0.3) y, n = 20 (10 male) or were adult, mean (SD) age = 24.9 (3.9) y, n = 20 (8 male). In pilot work with adults we could find fusion only with planes whose edges were visible; therefore we presented elliptical discs cut from planes similar to those in the first experiment (Fig. S2). On each trial a 45° standard disk was compared with a test disk of either the same slant or a slant differing by ±25° (for 6-y-olds) or ±12.5° (for adults) on the basis of single or combined cues (Fig. 2*A*). The stimuli are described in detail in *SI Materials and Methods*.

Observers first practiced making same/different slant judgments using the experimenter's hands. They then completed 30 trials (equal numbers of "same" and "different," randomly mixed) in each of six conditions. These comprised (Fig. 2*A*) combined-cue conditions *T+D+* and *T+D−* and single-cue conditions *D+*, *D−*, *T+*, and *T−* (the latter was included so that conditions could be presented in pairs with unpredictable directions of slant, but was not analyzed). Trials from pairs of conditions *T+* and *T−*, *D+* and *D−*, and *T+D+* and *T+D−* were mixed in blocks of 10 comprising 5 of each; the order in which these three condition pairs were presented was random. Conditions were mixed so that observers would not learn the strategy of looking for one direction of slant difference during a block. This strategy could artificially lower performance on fusion condition *T+D−* that includes both directions. There were 180 total trials with an opportunity for a pause every 10. A discrimination score (d') (19) was calculated for each condition as the z-score of the hit rate minus the z-score of the false alarm rate. Combined-cue condition scores were predicted from single-cue condition scores by decision models using integration and probability summation (*SI Materials and Methods*). Participants with d' scores ≤0 on more than one condition were excluded from analysis (n = 4; two children, two adults); in all four excluded participants the ≤0 scores were in single-cue conditions.

Experiment 3: Latency. Participants, all showing stereo vision on the TNO test (33), were aged 6, mean (SD) age = 6.5 (0.3) y, n = 15 (8 male) or adult, mean (SD) age = 23.9 (3.6) y, n = 17 (9 male). Stimuli were similar to those from experiment 1 (*SI Materials and Methods*). The order of the three conditions (*D*, *T*, *DT*) was random. Each condition began with practice trials comparing the 45° standard with a 70° comparison. After four consecutive successful judgments, testing began in sets of 10 trials during which observers pressed the left or the right arrow key to judge the more slanted plane. Within a condition, trials were shuffled in 20 blocks comprising all levels: thus for 6-y-olds 20 sets of 3 (5°, 12.5°, 35°) and for adults 20 sets of 7 (0°, 2.5°, 5°, 7°, 12.5°, 20°, 35°). To make results comparable with the integration experiment, we did not emphasize speed, but only encouraged participants to "concentrate" for each set of 10 trials without talking. Under these circumstances some participants were distractable, with long response latencies due to inattention to the task. We excluded from analysis those with 5% of responses >5 s (3 children, 2 adults) and one child who was at chance on stereo judgments. We excluded any trials that were "extreme outliers" (generally due to inattention to the task, e.g., stopping to ask the experimenter a question), defined as latencies ≥3× the third quartile + the interquartile range of latencies for each observer (36). On this criterion, 2.4% of 6-y-old and 1.2% of adult trials were excluded. The remaining responses for each level and condition were averaged to give an observer's proportion correct and mean response latency. The estimate of 6-y-olds' lapse rate related to interpretation of results from experiment 1 was made using data from all trials irrespective of their latency, i.e., before excluding any trials or participants.

ACKNOWLEDGMENTS. We thank Meera Desai for help with testing and Dorothy Cowie for useful discussions. This work was funded by United Kingdom Economic and Social Research Council Grant RES-062-23-0819 and by a Wellcome Trust vacation scholarship to Meera Desai.

1. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433.
2. Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.
3. Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: Optimal cue combination. *J Vis* 4:1–24.
4. Clark JJ, Yuille AL (1990) *Data Fusion for Sensory Information Systems* (Kluwer Academic, Boston).
5. Yuille AL, Bulthoff HH (1996) Bayesian theory and psychophysics. *Perception as Bayesian Inference*, eds Knill DC, Richards W (Cambridge Univ Press, New York), pp 123–162.
6. Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. *Science* 218:1138–1141.
7. Bahrick LE, Lickliter R (2004) Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cogn Affect Behav Neurosci* 4:137–147.

8. Sann C, Streri A (2007) Perception of object shape and texture in human newborns: Evidence from cross-modal transfer tasks. *Dev Sci* 10:399–410.
9. Nardini M, Jones P, Bedford R, Braddick O (2008) Development of cue integration in human navigation. *Curr Biol* 18:689–693.
10. Gori M, Del Viva M, Sandini G, Burr DC (2008) Young children do not integrate visual and haptic form information. *Curr Biol* 18:694–698.
11. Barutchu A, Crewther DP, Crewther SG (2009) The race that precedes coactivation: Development of multisensory facilitation in children. *Dev Sci* 12:464–473.
12. Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: Mandatory fusion within, but not between, senses. *Science* 298:1627–1630.
13. Ernst MO (2005) A Bayesian view on multimodal cue integration. *Human Body Perception from the Inside Out*, eds Knoblich G, Thornton IM, Grosjean M, Shiffrar M (Oxford Univ Press, New York), pp 105–131.
14. Gori M, Sandini G, Martinoli C, Burr D (2010) Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr Biol* 20: 223–225.
15. Knill DC (1998) Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Res* 38:1655–1682.
16. Rogers BJ, Bradshaw MF (1995) Disparity scaling and the perception of frontoparallel surfaces. *Perception* 24:155–179.
17. Blake A, Bülthoff HH, Sheinberg D (1993) Shape from texture: Ideal observers and human psychophysics. *Vision Res* 33:1723–1737.
18. Knill DC (1998) Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Res* 38:1683–1711.
19. Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics* (Wiley, New York).
20. Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
21. Raab DH (1962) Statistical facilitation of simple reaction times. *Trans N Y Acad Sci* 24: 574–590.
22. Laming DRJ (1968) *Information Theory of Choice-Reaction Times* (Academic, New York).
23. Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. *Psychol Sci* 9:347–356.
24. Gu Y, Angelaki DE, Deangelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11:1201–1210.
25. Beck JM, et al. (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60:1142–1152.
26. Welchman AE, Deubelius A, Conrad V, Bülthoff HH, Kourtzi Z (2005) 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci* 8: 820–827.
27. Iliescu B, Dannemiller JL (2008) Brain-behavior relationships in early visual development. *Handbook of Developmental Cognitive Neuroscience*, eds Nelson CA, Luciana M (MIT Press, Cambridge, MA), 2nd Ed, pp 127–146.
28. Schorr EA, Fox NA, van Wassenhove V, Knudsen EI (2005) Auditory-visual fusion in speech perception in children with cochlear implants. *Proc Natl Acad Sci USA* 102: 18748–18750.
29. Putzar L, Goerendt I, Lange K, Rösler F, Röder B (2007) Early visual deprivation impairs multisensory interactions in humans. *Nat Neurosci* 10:1243–1245.
30. Ballard DH, Hayhoe MM, Pook PK, Rao RP (1997) Deictic codes for the embodiment of cognition. *Behav Brain Sci* 20:723–742; discussion 743–767.
31. Kail R (1991) Developmental change in speed of processing during childhood and adolescence. *Psychol Bull* 109:490–501.
32. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
33. Cooper J, Feldman J, Medlin D (1979) Comparing stereoscopic performance of children using the Titmus, TNO, and Randot stereo tests. *J Am Optom Assoc* 50: 821–825.
34. Kaernbach C (1991) Simple adaptive testing with the weighted up-down method. *Percept Psychophys* 49:227–229.
35. Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Res* 35:389–412.
36. Tukey JW (1977) *Exploratory Data Analysis* (Addison–Wesley, Reading, MA).

Supporting Information

Nardini et al. 10.1073/pnas.1001699107

SI Materials and Methods

Experiment 1: Cue Integration. Each plane was projected as if seen through a 18.5 × 25-cm aperture in a gray screen at the same distance as the monitor screen; thus the planes were projected as if they were behind the monitor. Planes' edges could not be seen. Observers' eyes, and the horizontal axis about which planes rotated, were aligned with the vertical center of the monitor. Planes' distances in depth were +25 and +25 ± 0–10 cm from the screen. Disparity and/or texture cues (Fig. 1*A*) signaled slants about the horizontal axis, away from the observer. The binocular cue to slant came from the gradient of interocular disparities. Because slants were about the horizontal axis, horizontal and vertical size ratios (1) were not useful; and as stimuli were projected on a flat screen at a fixed distance, vergence and accommodation (1, 2) were not useful.

Planes were first composed of a regular grid of points with average density 0.6 points/cm (±0–10% for each plane on each trial) in both *x* and *y* directions. Each point was then randomly jittered in *x* and *y* directions by ±0–0.3 grid squares for conditions *T* and *DT* or ±0–1.5 grid squares for texture-absent condition *D* (effectively producing a random spread; Fig. 1*A*). In conditions *T* and *DT* each point formed the center of a randomly colored tile whose vertices were defined by a Voronoi tessellation (3). In condition *D* the points themselves were projected as randomly colored dots 0.2 cm wide with appropriate disparities signaling slant but without the changes in dot size and density that would provide a texture cue to slant (see below). Tiles and dots had anti-aliased black outlines 1.5 pixels wide.

So that disparity and texture information could be dissociated, each dot or vertex was first projected with respect to a cyclopean eye on the basis of its texture-defined slant. A new spatial location for the point was calculated, corresponding to where the point would be if its cyclopean projection had come from a plane with the disparity-defined slant. This point was then projected to the two eyes with the interocular disparity appropriate for its new distance. Thus in condition *D* dots had slant 0° from the vertical signaled by texture, but slant ≥45° signaled by disparity. In "high slant" and "low slant" conflict conditions, tiled planes with equal slants from texture and disparity were compared with planes with 10° mismatches in these cues.

To estimate a 75% threshold for each observer in conditions *D*, *T*, and *DT*, we used least-squares fits of psychometric functions to the collected data (all responses following practice). For condition *D* a cumulative Gaussian was fitted, predicting probability of a correct response as a function of slant difference, with mean 0 and variance a free parameter. For conditions *T* and *DT* that include texture information, a cumulative Gaussian was fitted with mean fixed at 0 and SD σ changing with slant according to an exponent dependent on a second free parameter, β (Eq. **S1**, where *N* is the cumulative normal distribution with mean μ = 0). This parameter allows for functions in which the variance of texture judgments decreases exponentially with increasing slant (4, 5). To prevent spurious fits of noisy data β was constrained between 0 and −0.05.

$$p\left(x\right) = N\left[x, \mu, \left(\sigma e^{\beta x}\right)^2\right]. \quad \textbf{[S1]}$$

We measured the point of subjective equality (PSE) in low slant and high slant conditions by fitting a cumulative Gaussian predicting probability of judging the comparison plane as the more slanted as a function of its slant. Mean and variance were free parameters. The PSE is estimated by the mean, i.e., the level at which the probability of judging either plane more slanted is 0.5. PSEs estimated beyond the bounds ±5° (at which the two planes could not be equal with respect to either cue) were estimated as + or −5°. In Fig. 1*B* PSEs are expressed as relative weights for texture, where weight 0 corresponds to a PSE of +5° (judging planes equal on the basis of disparity) and 1 corresponds to a PSE of −5° (judging planes equal on the basis of texture).

The ideal observer prediction for condition *DT* was calculated as follows. For two sensory information sources whose variances σ^2_A and σ^2_B are fixed, the predicted combined variance σ^2_{AB} given integration using optimal weights is (6, 7)

$$\sigma^2_{AB} = \frac{\sigma^2_A \sigma^2_B}{\sigma^2_A + \sigma^2_B}. \quad \textbf{[S2]}$$

We modeled variability as determined by a single parameter σ_D for disparity estimates and by two parameters, σ_T and β, which are scaled by slant *x*, for texture estimates (Eq. **S1**). The optimal predicted variance of estimates σ^2_{DT} at slant *x* is then

$$\sigma^2_{DT}\left(x\right) = \frac{\left(\sigma_D e^{\beta x}\right)^2 \sigma^2_T}{\left(\sigma_D e^{\beta x}\right)^2 + \sigma^2_T}. \quad \textbf{[S3]}$$

The proportion of correct responses at each level of slant *x* was predicted by a cumulative Gaussian with variance σ^2_{DT} (*x*) changing as a function of slant *x* (Eq. **S3**). The predicted optimal threshold was the level of slant predicting 75% correct responses.

The rationale for the prediction that texture will be more reliable for near-horizontal than near-vertical planes comes from refs. 4 and 8 and is illustrated in Fig. S1. Rotation of a plane to increase its slant reduces the height of a single tile's 2D projection (Fig. S1 *A* and *B*). However, the same degree of rotation (slant increase) has smaller effects on the tile's projection when the plane is near orthogonal (Fig. S1*A*) than when it is near parallel (Fig. S1*B*) to the line of sight. Consistently with this, a 10° slant difference is harder to detect near the vertical (Fig. S1*C*) than near the horizontal (Fig. S1*D*). Because detecting slant differences from texture depends on detecting differences in tiles' projections (e.g., in their horizontal-to-vertical size ratios) (4, 8), and differences in projection are largest for planes near the horizontal, texture provides increasingly useful slant information as the plane approaches the horizontal. Therefore, we predict that observers using texture cues to slant will find these cues more useful as the plane approaches the horizontal. Although we cannot assume a linear (or any other) relationship between the magnitudes of 2D projection differences and observers' sensitivity to them, we can assume that the relationship is monotonic in the predicted direction.

Experiment 2: Fusion. Planes were constructed as before, with regular grids of 1 point/cm randomly jittered by ±0–0.225 cm when the texture cue was present or ±0–1.5 cm for disparity-only (dots) conditions. Dots were sized 0.35 cm. Each disk was an elliptical region of tiles selected from these planes. Discs' centers were projected at depth +10 or −10 cm relative to the screen. Example stimuli are shown in Fig. S2. Six-year-olds' disparity-based same/different judgments for these discs were poor; to increase the usefulness of disparity we therefore tested children at 87.5 cm, half the adult viewing distance. Each disk had width 16 cm and a randomly chosen depth between 1 and 1.5 times its width (for adults) or between 0.5 and 1.5 times its width (for 6-y-olds). The larger range of aspect ratios for 6-y-olds prevented the vertical size of the disk's projection becoming a fully reliable cue to slant, given the greater slant differences shown to this group (±25° vs. ±12.5° for adults). At both ages, in conditions with texture signaling greater (+) slant, the vertically smaller disk was the more slanted in an average 93% of trials.

Performance in the fusion experiment was compared with an integration model using combined disparity and texture cues and with two probability summation models using single cues. These models are formulated using signal detection theory (9, 10). The sensory estimates observers use to categorize pairs of discs as being of *same* or *different* slant are modeled as coming from overlapping normal distributions with unit SD. The distribution of slant differences perceived when viewing discs with the *same* slant is centered on 0 (no difference). The distribution of differences perceived when viewing discs with *different* slant is centered on d', which therefore expresses the distance between the two distributions' peaks in units of their common SD. d' is estimated by the z-score of the observer's hit rate minus the z-score of the false alarm rate (9, 10). The models use estimates of d' from each observer's single-cue conditions to predict how the observer will perform given both cues at the same time. In the integration model, decisions are made with respect to a combined estimate of slant based on both disparity and texture information. In the probability summation models, independent decisions are reached on the basis of the two single cues, and a slant difference is detected if either one or both single cues indicate a difference.

Integration Model. Integration of disparity and texture information into a joint metric of slant predicts an increase in combined-cue d' relative to both single cues when cues are congruent (condition $T{+}D{+}$). If the distances between 0 and d' in each single dimension (d'_{T+}, d'_{D+}) are described by two orthogonal lines in a 2D space, the distance between 0 and the joint d' in two dimensions (d'_{T+D+}) is given by completing the right-angled triangle via Pythagoras' theorem (Fig. S3 and ref. 10):

$$d'_{T+D+} = \sqrt{{d'_{T+}}^2 + {d'_{D+}}^2}. \qquad \textbf{[S4]}$$

To achieve optimal discrimination, i.e., one benefiting from the separation between the two distributions given by the d' value from Eq. **S4** (Fig. S3*A*, dotted blue lines), a decision criterion is required that is perpendicular to the (0, 0)–(d'_{D+}, d'_{T+}) line and cuts diagonally across the 2D space (Fig. S3*A*, dashed green line). This reprojects the 2D distributions onto a single "decision axis" (10) expressing the slant difference given by the sum of disparity and texture differences (Fig. S3*A*). In other words, under this model decisions are based on the sum of disparity-based and slant-based estimates. Assuming that these estimates are independent, the resulting reduction in noise results in better discrimination (higher d') for congruent stimuli.

However, decisions along this same axis with respect to conflicting stimuli (condition $T{+}D{-}$) have very low d', because *same* and *different* stimuli project onto similar places on the decision axis (Fig. S3*A*; the distribution for *different* stimuli in condition $T{+}D{-}$ would be centered at d'_{D-}, d'_{T+} and its projection on the decision axis would be the very close to that for *same* stimuli). In the special case in which the three values d'_{D+}, d'_{D-}, and d'_{T+} are equal, the distributions for *same* and conflicting *different* stimuli overlap exactly on the decision axis, predicting a d' score of 0 (total inability to make the discrimination). When these values are unequal (as was the case with most individual observers), and in general, the predicted d' is

$$d'_{T+D-} = -\sqrt{{d'_{D-}}^2 + {d'_{T+}}^2}\cos\left(\tan^{-1}\left(\frac{d'_{T+}}{d'_{D+}}\right) + \tan^{-1}\left(\frac{d'_{T+}}{d'_{D-}}\right)\right). \qquad \textbf{[S5]}$$

Because observers' single-cue d' values were not exactly equal, for both 6-y-olds and adults the mean fusion prediction was for scores slightly better than 0 (Fig. 2*B*).

Probability Summation Model 1 (Liberal Criterion). In the probability summation models (Fig. S3*B*), same/different decisions are reached independently on the basis of disparity and texture. In the first ("liberal") version of the model, a difference between stimuli is detected if *either* decision criterion is reached (Fig. S3*B*, all shaded areas). Combined-cue hit and false alarm rates are first calculated by multiplying single-cue hit and false alarm rates (10),

$$H_{D+T+} = H_{D+} + (1 - H_{D+})\, H_{T+} \qquad \textbf{[S6]}$$

$$F_{D+T+} = F_{D+} + (1 - F_{D+})\, F_{T+}, \qquad \textbf{[S7]}$$

where H is hit rate and F is false alarm rate. The combined-cue d' is then calculated as the z-score of the combined hit rate minus the z-score of the combined false alarm rate. The calculations for conflicting condition $T{+}D{-}$ are similarly based on hit and false alarm rates for the component cues.

Probability Summation Model 2 (Conservative Criterion). In the "conservative" version of the model, a difference between stimuli is detected only if *both* single cues indicate a difference (Fig. S3, darker shaded area only). Combined-cue hit and false alarm rates are calculated by multiplying single-cue hit and false alarm rates (10) ($H_{D+T+} = H_{D+}H_{T+}$, $F_{D+T+} = F_{D+}F_{T+}$, calculating rates for conflicting condition $T{+}D{-}$ from its component cues in the same way).

Unlike the integration model, the probability summation models do not integrate the cues into a common metric. Therefore, there is nothing special about a conflicting stimulus making it in principle harder to detect than a congruent stimulus. However, under the probability summation models potential improvements in d' relative to single cues are modest, and decreases in d' relative to the best single cue are also possible.

Experiment 3: Latency. Stimuli were the same as those from the first study (Fig. 1*A*), except that the screen through which stimuli were seen was black, dots were sized 0.35 cm, and planes' centers' depths relative to the screen were +45 ± 10 cm.

1. Backus BT, Banks MS, van Ee R, Crowell JA (1999) Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Res* 39:1143–1170.
2. Rogers BJ, Bradshaw MF (1995) Disparity scaling and the perception of frontoparallel surfaces. *Perception* 24:155–179.
3. de Berg M, can Kreveld M, Overmars M, Schwarzkopf O (2000) *Computational Geometry: Algorithms and Applications* (Springer, New York), 2nd Ed.
4. Knill DC (1998) Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Res* 38:1655–1682.
5. Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: Optimal cue combination. *J Vis* 4:1–24.
6. Clark JJ, Yuille AL (1990) *Data Fusion for Sensory Information Systems* (Kluwer Academic, Boston).
7. Ernst MO (2005) A Bayesian view on multimodal cue integration. *Human Body Perception from the Inside Out*, eds Knoblich G, Thornton IM, Grosjean M, Shiffrar M (Oxford Univ Press, New York), pp 105–131.
8. Blake A, Bülthoff HH, Sheinberg D (1993) Shape from texture: Ideal observers and human psychophysics. *Vision Res* 33:1723–1737.
9. Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics* (Wiley, New York).
10. Macmillan NA, Creelman CD (2004) *Detection Theory: A User's Guide* (Lawrence Erlbaum, New York), 2nd Ed.

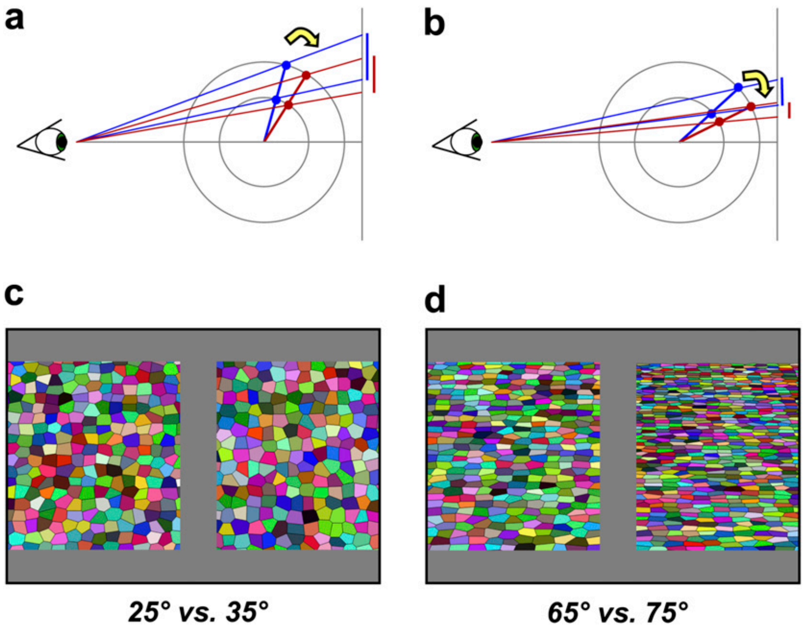

Fig. S1. Projection of textured planes with low and high slant. (*A* and *B*) Effects of a change in slant on the projected vertical height of a single tile at low slant (*A*) and high slant (*B*). (*C* and *D*) Stimuli with a 10° difference in slant at low slant (*C*; the *Left* is more slanted) and high slant (*D*; the *Right* is more slanted).

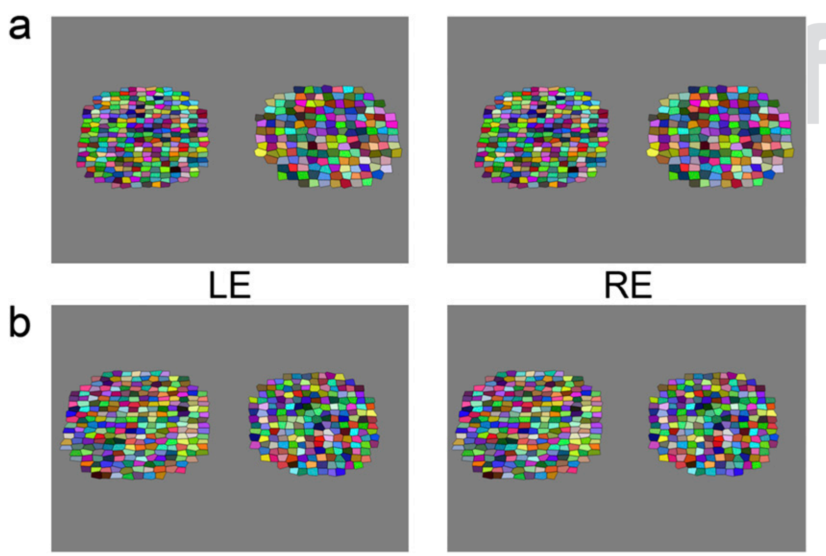

Fig. S2. Stimuli for fusion experiment. Left eye (LE) and right eye (RE) views of planes with 12.5° slant differences (*A*) in the same direction in terms of both disparity and texture (condition *T*+*D*+) and (*B*) in opposite directions (condition *T*+*D*−). To view stereoscopically, free fuse by diverging the eyes (slants from disparity may differ from those in the experiment, where these images took up 13° of visual angle).

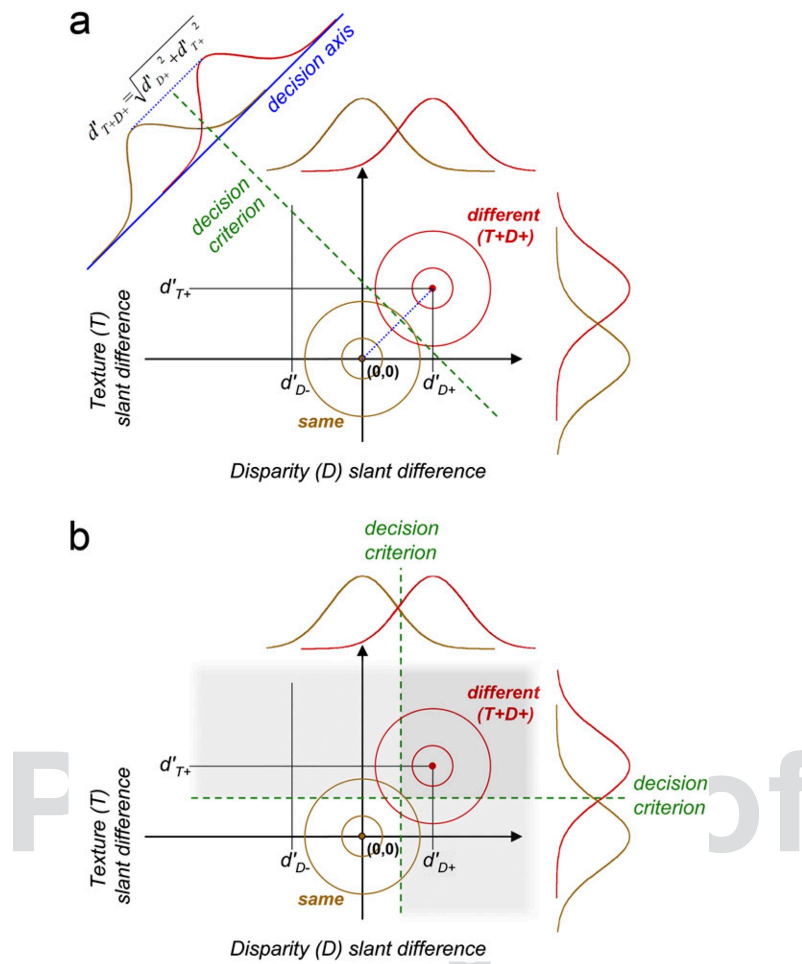

Fig. S3. Models for fusion experiment. (*A*) Integration model, in which decisions are based on the sum of disparity and texture estimates of slant difference. (*B*) Probability summation model, in which independent decisions are reached for each component cue. A difference is detected if either single cue indicates a difference (liberal criterion, all shaded areas) or if both cues indicate a difference (conservative criterion, dark shaded area only).