Bayesian modeling of cue interaction: bistability in stereoscopic slant perception

Raymond van Ee

Helmholtz Institute, Utrecht University, PrincetonPlein 5, 3584CC Utrecht, The Netherlands

Wendy J. Adams and Pascal Mamassian

Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland

Our two eyes receive different views of a visual scene, and the resulting binocular disparities enable us to reconstruct its three-dimensional layout. However, the visual environment is also rich in monocular depth cues. We examined the resulting percept when observers view a scene in which there are large conflicts between the surface slant signaled by binocular disparities and the slant signaled by monocular perspective. For a range of disparity–perspective cue conflicts, many observers experience bistability: They are able to perceive two distinct slants and to flip between the two percepts in a controlled way. We present a Bayesian model that describes the quantitative aspects of perceived slant on the basis of the likelihoods of both perspective and disparity slant information combined with prior assumptions about the shape and orientation of objects in the scene. Our Bayesian approach can be regarded as an overarching framework that allows researchers to study all cue integration aspects—including perceptual decisions—in a unified manner. © 2003 Optical Society of America

OCIS codes: 330.0330, 330.1400, 330.4060, 330.7310, 330.5510.

1. INTRODUCTION

A task of the visual system is to infer the scene that best explains our incoming retinal information. This is not a straightforward task because our three-dimensional (3D) observations depend not only on the retinal images but also on the brain's assumptions about the world.[1] One instance in vision in which a given retinal image produces a changing 3D observation within an unchanging stimulus is the phenomenon of perceptual bistability. Perceptual bistability is an interesting phenomenon because it creates the opportunity of having two states in neural processing that are modulated by the brain's assumptions about the world rather than the stimulus.

Here we study perceptual bistability by making use of the distinct binocular and monocular depth information in an image. In binocular viewing, binocular disparities arise because our eyes view a scene from slightly different positions. These disparities enable us to reconstruct the 3D layout. The processing of disparities is, however, not essential for the 3D reconstruction; monocular cues can be sufficient to recover 3D structure. For example, linear perspective is a powerful cue for surface orientation. The integration of perspective and disparity has been the subject of quite a few studies.[3–12] However, the bistability that can be created when the monocular and binocular cues in a scene specify opposite depth information has received only little interest in the scientific literature,[13–19] and no studies have modeled the quantitative aspects of this phenomenon.

To evoke bistability, observers viewed ambiguous stereoscopic images of a grid stimulus in which disparity and perspective specified different slants.[24] This stimulus is similar to Ames's famous trapezoid window. Slant here refers to rotation about a vertical axis through the center of the stimulus. The grid was viewed against a surrounding frontoparallel reference surface. Figure 1 shows an example of the stimulus used in this study. On inspection, the reader might be able to distinguish the two 3D percepts that are present when linear perspective and binocular disparity specify opposite slants: One percept in which the grid's slant is near to the disparity-specified slant, and the other in which the perceived slant is closer to the perspective-specified slant. The two percepts were never present simultaneously. Most observers have no difficulty in reporting two distinct slant percepts when presented with stimuli containing large cue conflicts.

In the following, we describe the results of a psychophysical experiment in which we presented stimuli with highly dissonant disparity and perspective information. We then present our Bayesian model—the main focus of this paper—which quantitatively accounts for individual observers' data.

2. METHODS

The stimuli (Fig. 1) have been described in detail in a previous paper.[24] In short, the stimuli were planar grids (subtending 15 deg × 11 deg in unslanted conditions) presented dichoptically by a conventional red–green anaglyphic technique. A surrounding pattern (92 deg × 39 deg) consisting of small squares (1 deg × 1 deg) provided a zero-slant reference. The stimuli were viewed on a large (92 deg × 77 deg) projection screen at a viewing distance of 114 cm.
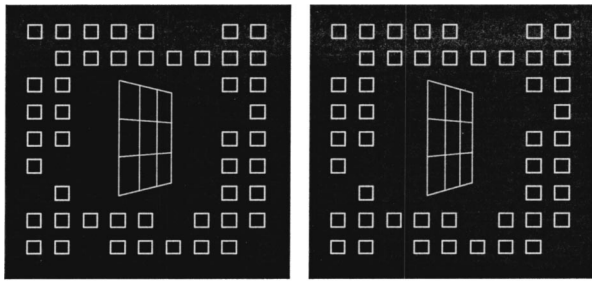
Fig. 1. In this stereogram both perspective and binocular disparity specify surface slant about the vertical axis. In uncrossed fusion of the stereogram (the left eye views the left image, and the right eye views the right image), two stable percepts can be distinguished. In the first percept the grid recedes in depth with its right side farther away (it is perceived as a slanted rectangle). In the other percept the left side of the grid is farther away (it is perceived as a trapezoid with the near edge shorter than the far edge). In fact, the slants depend on the viewing distance; however, the slant signs are conflicting regardless of the viewing distance. Note that each of the two percepts can be selected and maintained at will in a relatively controlled fashion. In crossed fusion, perspective and disparity specify similar slants and the observer perceives a single stable slanted rectangular grid with its right side further away.
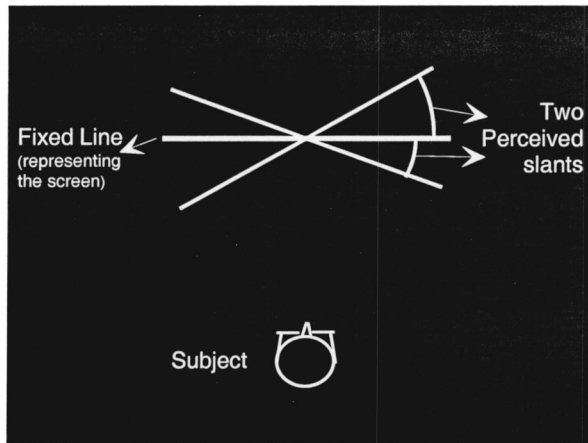

Fig. 2. Schematic drawing of the slant-estimation method representing a top view of the viewing geometry. One of the lines was fixed, and the other two lines could be rotated about their centers. The fixed line represented zero slant (the image plane on the frontal screen); each of the other lines represented the perceived grid in either the perspective-dominated percept or in the disparity-dominated percept. Using this display, subjects matched the perceived slant(s) to the angle(s) between the fixed horizontal line and the rotatable intersecting line(s).

A subject initiated the stimulus onset by a mouse click. The presentation duration was 10 s, and the subject then estimated the perceived slant of the grid. The subjects were instructed that both ambiguous (flip) and nonambiguous (nonflip) stimuli would be presented. They were also informed that the stimuli could be either trapezoidal or rectangular. Subjects were free to move their eyes.[25] Eight subjects participated. All had normal or corrected-to-normal vision and completed a stereoanomaly test.[27] Subjects were naïve to the purpose of the experiment.

The slant-estimation procedure[28] is depicted in Fig. 2. To make the slant estimates, three frontoparallel lines were presented on the screen after the stimulus presentation. One of the lines was horizontal, and the other two lines could be rotated about their centers. The hori-

zontal line was fixed and represented a top view of the unslanted reference; each of the other lines represented the top view of the perceived grid in either the perspective-dominated percept or in the disparity-dominated percept. Subjects were instructed to match the angles between the rotatable lines and the horizontal line to the two perceived slants. If an observer was not able to experience bistability, he or she matched both angles to the (single) slant he or she perceived. Because the lines were displayed stereoscopically in the plane of the screen, they also served as a zero-slant reference between successive stimuli.[29]

To investigate systematically how perspective and disparity information contribute to bistable 3D perception, we varied both disparity-specified slant (-70 to 70 deg in ten steps) and perspective-specified slant (-70 to 70 in six steps). Positive slants are defined as right side away. In each block of 77 trials, all the stimulus conditions appeared once in random order. There were five trial blocks.

3. RESULTS

Figure 3 shows the individual data for two subjects. The plots depict the mean perceived slants across the five trial repetitions for a range of perspective and disparity slants. The data for each of the subjects in Fig. 3 can be roughly split into two domains. In one domain, when disparity and perspective specified similar slants, then only one perceived slant was reported. In this domain, slants derived from perspective and disparity have been reconciled; a perceived slant is produced somewhere between the two. In the other domain, when disparity and perspective specified quite different slants, subjects experienced bistability and reported two perceived slants. Data have been reported for a similar experiment in a previous paper.[24] Although the study did not report individual subject data on bistability, the two sets of data followed very similar patterns. In the previous study bistability occurred when the perspective- and the disparity-specified slants had opposite signs. In the current study, bistability also occurred when the two had the same sign. In the current study, subjects were informed that the stimuli could be either trapezoidal or rectangular. In the previous study, subjects were merely asked to report bistability when they were able to perceive one plane with its left side in front and another plane with its right side in front. This difference in instruction could account for the observers' slightly different behaviors, although large interobserver differences were observed in both studies.

It is interesting to consider the observer's two percepts in the conditions under which bistability occurred. In this domain, observers commented that at one of the reported slants the object appeared trapezoidal and at the other reported slant the object appeared rectangular. Up to this point we have been referring to our stimuli as "cue conflict." However, strictly speaking, our stimuli do not present a conflict between information sources. All stimuli are consistent with a real-world object, which may be trapezoidal or rectangular. The stimuli present a conflict only under the assumption that the original object is
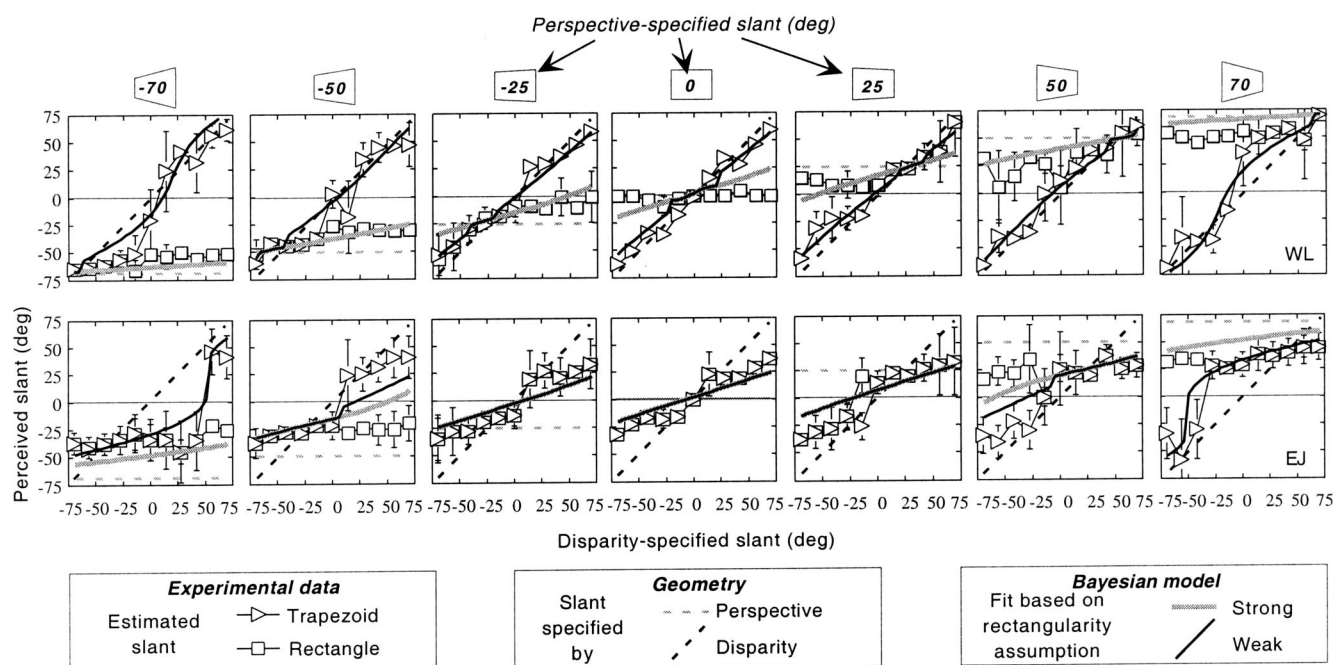
Fig. 3. Perceived slant and the Bayesian fits as a function of disparity-specified slant for a range of different perspective-specified slants. Each row of panels represents the data of one subject. The top row depicts the best fit that accounted for 93% of the variance in the data. The bottom row depicts the worst fit that accounted for 79% of the variance in the data. The fits to the data produced by the Bayesian model are indicated by the gray and black curves. The gray curves indicate the strong rectangular assumption, and the black curves indicate the weak rectangular assumption. The subjects perceived either a slanted rectangular grid (square symbols) or a slanted trapezoid (triangles). The slants that were geometrically present in the stimulus are represented by the dashed prediction lines. Error bars represent ± 1 standard deviation in the mean across the five trial repetitions.

approximately rectangular. It is only by using this type of assumption that linear perspective can be informative. We assume that observers flipped between the two perceived slants by changing the strength of this rectangularity assumption, and this forms the basis of our model.

4. BAYESIAN MODEL

To understand the cue interaction that engenders bistability in stereoscopic slant perception, we developed a Bayesian model. A Bayesian model combines multiple sources of information in an optimal way with the ultimate goal of maximizing performance in a particular task.[32] Bayesian modeling has been successfully applied in computer vision (see the review by Knill and Richards[41]), and, in the past decade, several investigators have started to apply this framework to human vision.[36,38,42–54]

In the present paper the multiple sources of information to be combined are the stereoscopic cue for slant, the perspective cue for slant, and a preference (prior) for frontoparallel. This prior distribution also encompasses the residual flatness cues in the display (for example, accommodative blur, the fixed graininess of the pixels on the screen, or the brightness gradient). As described in more detail in the next paragraphs, the perspective and disparity likelihoods can be computed directly from the images, with some basic assumptions about how tolerant to noise the visual system is. Once combined, likelihood and prior provide a posterior probability function that assigns a probability to each possible event in the world, in our case, each possible slant of the surface. The last stage of

a Bayesian model is a decision rule that translates this posterior probability into an actual response. Details of the model are provided in Appendix A.

Linear perspective information in an image can be exploited only by making assumptions about the orientations of the contours in the world that are projected onto the image plane. In our model, perspective information is interpreted by assuming that the object is roughly rectangular. This rectangularity assumption is implemented by assuming that the orientation of (nonvertical) lines in the image plane can be described by a Gaussian distribution centered on zero (horizontal). The width of this Gaussian is a free parameter in our model and reflects the strength of the rectangularity assumption. The orientation of a line in the image can easily be computed from the orientation of the line on the object, the surface slant, and other scene parameters (see Fig. 4 and Appendix A). We assume that any error in measuring the image line orientations is negligible in the context of a robust rectangularity assumption. Given these assumptions, we can calculate the probability, for any surface slant, of getting lines with the orientations and elevations measured in the image. This distribution of probabilities of the image, given the surface slant $p(I|S)$, is the perspective likelihood. By calculating the perspective likelihood in this way, we are reflecting the amount of information present in the image at different slants. At large slants, perspective information is more reliable, and our model incorporates this (Fig. 5).

The likelihood for the disparity information is modeled more simply. The disparity likelihood is here modeled as a Gaussian centered on the true slant of the surface. The

width of the Gaussian reflects measurement noise and is the first parameter of the model.

The third distribution to consider is the prior. This is modeled by a Gaussian centered on zero. This reflects our prior assumption that objects in the world are close to frontoparallel, as well as incorporating all residual cues in the stimulus, such as accommodation, vergence, texture, and blur cues that are consistent with a frontoparallel surface. The spread of this distribution is the second parameter of the model.

The optimum way of combining all this information is multiplication. The product of the two likelihoods and the prior distributions gives the posterior after normalization (the posterior is a probability distribution function, whereas the likelihoods need not be probability distributions). From Bayes's theorem,[55] this is proportional to the probability of the surface slant, given the image, $p(S|I)$.

In the current experiment, observers are asked to flip between two percepts—a perspective-dominated slant and a disparity-dominated slant. We have modeled this by allowing the model to work in two modes. In the perspective-dominant mode, we assume that the observer is implementing a strong rectangularity constraint. In

Fig. 5. Normalized likelihood $p(I|S)$ for perspective information computed from the geometry of perspective projection, assuming that the perceived orientation of each of the horizontal lines projected onto the image is subject to noise. The noise is assumed to be Gaussian centered on zero and with a standard deviation. Each curve shows the likelihood for one of the seven perspective-specified slants. For larger perspective-specified slants, the likelihood distribution is more peaked, reflecting an increased certainty in determining surface slant.
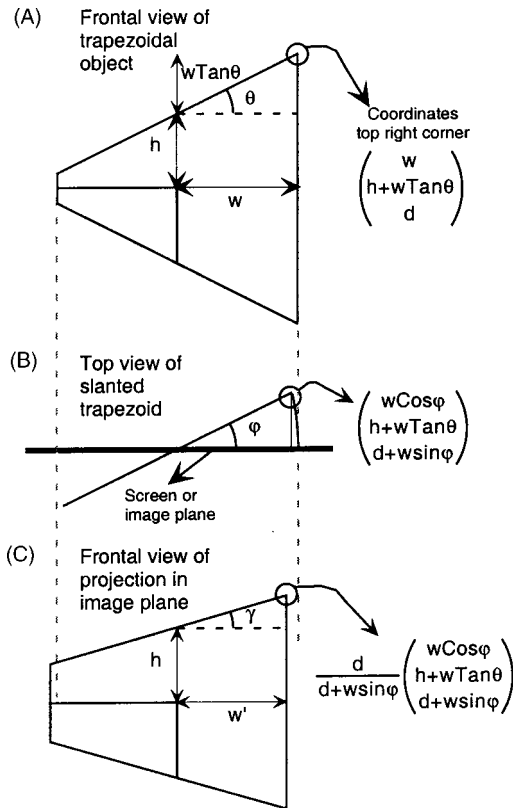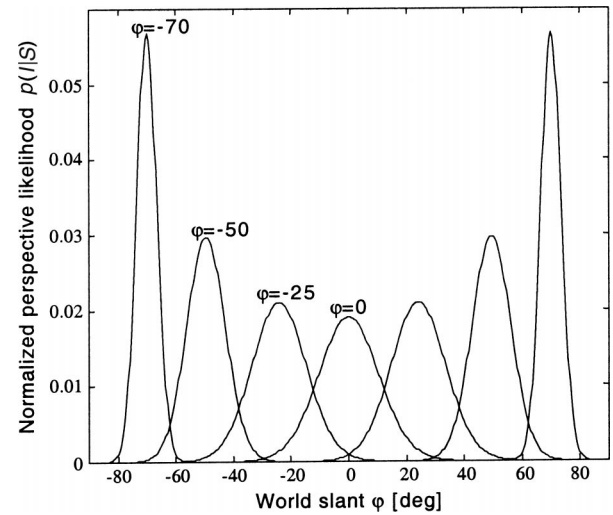

Fig. 4. Nonvertical stimulus lines change their orientation in the image plane when the stimulus is rotated about the vertical axis. (A) Frontal view of an unslanted trapezoidal object (w is the width, h is the central height, and d is the viewing distance). The orientation of the depicted stimulus line is θ. The 3D coordinates relative to the midpoint between the eyes are explicitly given. (B) A top view of the object after it has been rotated through an angle φ. (C) A projection of the slanted trapezoid on a frontal screen. The stimulus line whose orientation was originally θ is now projected with an orientation γ [see Eq. (A2) in Appendix A].

other words, he or she is assuming that the object in the world was a rectangle, and deviations from rectangularity in the image are a consequence of perspective projection. In the model, this translates into a smaller spread of the Gaussian describing the world line orientation (third parameter of the model).

In the disparity-dominant mode we assume that the observer is implementing a weak rectangularity assumption (a fourth parameter to characterize a wider distribution of line orientations). In other words, the original object can be a range of shapes. In effect, the influence of perspective becomes weaker (the likelihood becomes less peaked), and the disparity information becomes more dominant. The first and second parameters (standard deviation of disparity and prior) are kept constant for both of these modes. Each of these modes gives a separate posterior distribution as its output.

The model then has to decide what to do with the outputs of the two (perspective and disparity) modes. We apply a gain function to the sum of the two posterior distributions. A gain function is often used in Bayesian modeling as a smoothing function, which makes the model robust to local minima.

In our model the gain function is a Gaussian with a variable standard deviation (the fifth and final model parameter). The effect of convolving the gain function with the combined posterior is to produce an expected gain distribution with a single peak if the two posterior distributions were sufficiently similar and a distribution with two peaks if the two posterior distributions were disparate. The single peak corresponds to observers making a single response, whereas the double peak corresponds to the two percepts in the bistable stimuli.[59] Using this approach, we have modeled the bifurcation of responses that the observers give, arising from a single cue-conflict stimulus (see also Ref. 60).

We found the values of the parameters that provided the best fit to each individual observer's data. The best and the worst fits of the model are plotted in Fig. 3. The model provides an excellent fit (accounting for 93% of the variance) to the data of observer WL.[61] Even the model's worst fit (the one for observer EJ) provides a reasonable fit that accounted for 79% of the data. Figure 6 shows the model fits (and also the raw data) for two observers, AB and KM, who did not follow the data pattern exhibited

by the other eight observers. These subjects participated in a previous experiment.[24] We include their data in the current paper because they showed interesting behavior that formed a challenge for our Bayesian model. Observations of KM seem almost entirely dominated by perspective; all data lines are near horizontal, showing little effect of disparity. In contrast, observer AB shows almost no effect of linear perspective; the data are similar for all seven perspective conditions. As can be seen in Fig. 6,
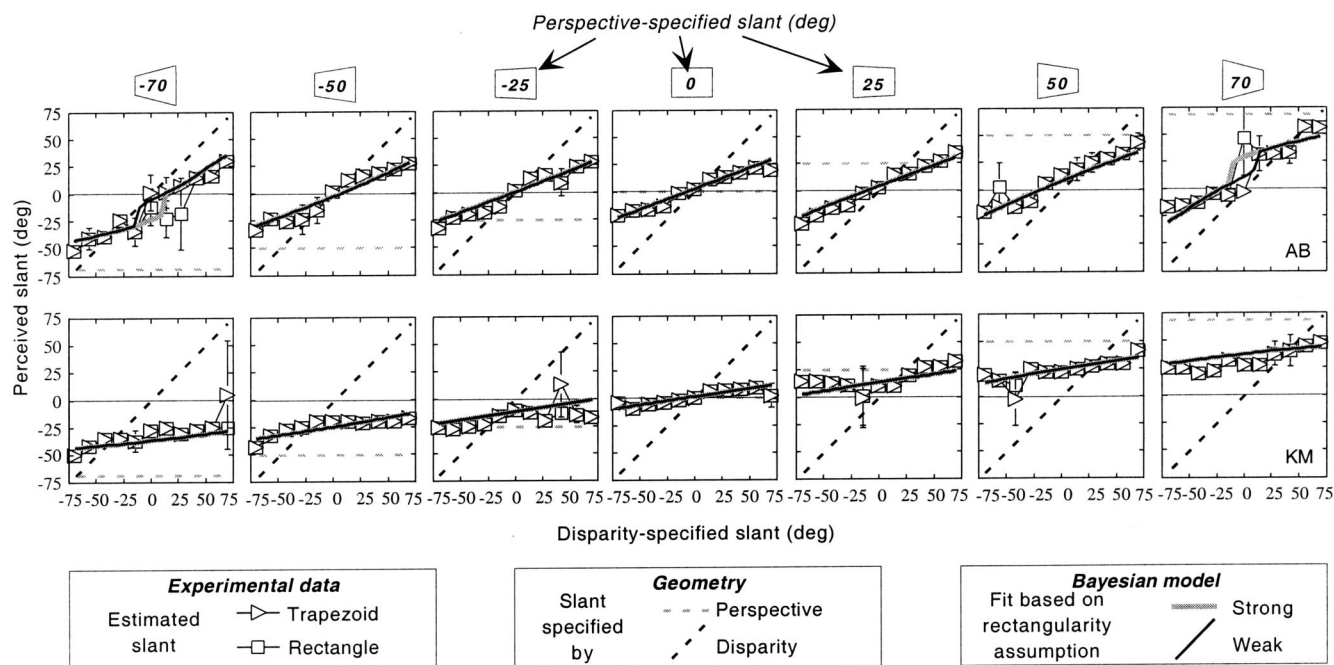

Fig. 6. Same as Fig. 3 but for two observers who did not follow the data pattern shown by the other eight observers. Both AB and KM observed almost no bistability. The data of AB are dominated by disparity information. The percept of observer KM seems almost entirely dominated by perspective; all data lines are near horizontal, showing little effect of disparity. Note that in almost all panels the fits produced by the Bayesian model fall on top of each other for both the weak and the strong rectangularity assumptions. The fits for AB and KM accounted for 92% and 93% of the variance in the data, respectively.

Table 1. Model Parameters for the Individual Observers[a]

| | Mode Dependent | | Mode Independent | | | |
| | σ of Perspective Likelihood for Rectangularity Assumption | | σ of disparity | SD[b] of | SD of | |
Subjects	Strong	Weak	Likelihood	Prior	Gain	R^2
WL (Fig. 3)	0.25	1.21	3.83	8.20	3.08	0.93
EJ (Fig. 3)	0.80	1.27	7.67	6.10	2.38	0.79
AB (Fig. 6)	3.03	7.18	17.35	14.30	1.65	0.92
KM (Fig. 6)	0.04	26.62	9.58	5.93	74.25	0.93
GE	0.40	0.85	5.16	3.86	3.91	0.81
EC	0.40	1.29	5.96	4.95	3.87	0.87
DL	0.31	1.44	4.46	5.93	0.29	0.88
BM	0.22	0.95	4.82	3.62	0.23	0.89
ER	0.61	0.98	5.02	4.14	0.77	0.90
BG	0.40	2.04	7.92	9.07	2.82	0.92

[a] The model operates in two modes; the perspective-likelihood parameter depends on whether strong or weak rectangularity of the stimulus is assumed. The other three parameters are mode independent. The last column contains the coefficient of determination R^2, an indication of the goodness of fit.
[b] Standard deviation.

the model provides a good fit for these two observers who were dominated by a single cue and did not report bistability.

Table 1 shows the parameters used in the model for all subjects. The last column of Table 1, the coefficient of determination R^2, gives an indication of the goodness of fit. The model accounts for 88% of the variance in the data on average (in the range of 79% to 93%). Given that there are 154 data points per observer (2 modes $\times$ 7 perspectives $\times$ 11 disparities) and only five parameters, we conclude that our Bayesian model accounts very well for the observers' performance in this task.

5. DISCUSSION

We have developed a Bayesian model for the quantitative aspects of bistability in perceived slant for a large spectrum of possible combinations of disparity- and perspective-specified slants. The model's account for the data is twofold. On the one hand, it accounts for the observation that subjects perceive only one slant when the perspective- and disparity-specified slants are similar. On the other hand, it accounts for the observation that subjects are able to select either a perspective- or a disparity-dominated slant when the specified orientations are rather different. The same model can also account for observers whose data follow a completely different pattern, by having a very noisy disparity signal (observer KM) or a very weak rectangularity constraint (observer AB).

The occurrence of a clear bifurcation has been reported before in an entirely different study[42] that used stereoscopic stimuli with mixed vertical disparity information consistent with two disparate gaze angles. However, this study did not report on perceptual bistability. Further, although quite a few other studies have used perspective and disparity cues that specified depth of opposite polarity,[3,4,8,42,43,62–68] none of these studies reported bistability. This could be because most of these studies have examined relatively small conflicts or short presentation times or both; in such circumstances observers might not notice bistability when they are not explicitly instructed to look for it.

That our Bayesian approach is rich enough to deal with both consonant and dissonant slant cues for a range of observers is an important feature that distinguishes it from existing models. It encompasses the integration of visual cues (1) in weak data fusion,[69,70] (2) in modified weak data fusion,[71,72] and (3) in strong fusion.[43] It is also an intuitive method by which to model Markov Random Fields[73] and the activity of neural populations.[74] A particular strength of the Bayesian approach, which differentiates it from other fusion models, is that it provides a natural way to include prior information with the information available in the image. In addition, it should be noted that the above-cited fusion models cannot account for both fusion and bistability without the addition of certain ad hoc robustness constraints. Our model can easily be extended to cues other than perspective and disparity. In other words, the Bayesian approach, as we developed it, can be considered an overarching framework in cue in-

teraction theory that allows researchers to study all above-mentioned cue integration aspects—including perceptual decisions—in a unified manner.

6. CONCLUSION

We have presented a coherent model of bistability in which each parameter has a clear and interesting meaning. There is one set of parameters (at the chosen viewing distance) that can explain perceptual bistability in stereoscopic vision for the complete spectrum of combinations of perspective and disparity.

APPENDIX A

We provide here details of the Bayesian model used to explain the cue interaction that engenders bistability in stereoscopic slant perception. The model combines the information from the perspective and disparity cues (the likelihood information) with a prior constraint. The outcome of this combination is then subjected to a decision rule.

A. Perspective Likelihood

Let θ be the orientation of a line on the surface before slanting that surface relative to the observer. From the rectangularity constraint, we assume that the orientation θ follows a Gaussian distribution centered on 0 (the horizontal orientation). The strength of the rectangularity assumption is left as a free parameter in the model and maps to the spread (standard deviation, r) of the Gaussian:

$$p(\theta; r) = \frac{1}{\sqrt{2\pi r^2}} \exp\left(\frac{-\theta^2}{2r^2}\right). \tag{A1}$$

From the geometry of the scene illustrated in Fig. 4, we can derive the relationship among the orientation of the line on the surface (θ), the projected orientation of the line in the image (γ), and the slant of the surface (φ):

$$\tan(\gamma) = \frac{d\tan(\theta) - h\sin(\varphi)}{d\cos(\varphi)}, \tag{A2}$$

where d is the viewing distance and h is the height of the line in the image. In this equation, the world slant (φ) is understood to span the range ($-\pi/2$, $\pi/2$), which is all the surface orientations between left slanted and right slanted.

From Eqs. (A1) and (A2), we can compute the likelihood $p(\gamma | \varphi; r)$, which is the probability of obtaining a particular line orientation given the slant of the surface. In practice, we simulated this likelihood by generating a large number of surfaces and storing in a matrix the obtained image line orientations.

B. Disparity Likelihood

The disparity likelihood is assumed here to be a Gaussian centered on the true disparity-specified slant (φ_d), with the spread (σ_d) of the distribution left as a free parameter:

$$p(d|\varphi; \sigma_d) = \frac{1}{\sqrt{2\pi\sigma_d{}^2}}\exp\left[\frac{-(\varphi - \varphi_d)^2}{2\sigma_d{}^2}\right]. \quad \text{(A3)}$$

C.   Prior Constraint
All the residual cues in the stimulus consistent with zero slant, together with a possible preference for frontoparallel surfaces, are modeled as a single prior constraint. This prior is assumed to follow a Gaussian distribution centered on zero slant and with the spread (σ_p) left as a free parameter:

$$p(\varphi; \sigma_p) = \frac{1}{\sqrt{2\pi\sigma_p{}^2}}\exp\left(\frac{-\varphi^2}{2\sigma_p{}^2}\right). \quad \text{(A4)}$$

D.   Combination of the Likelihoods and Prior
Likelihoods and prior are combined to produce the posterior probability. In general, the posterior represents the probability that a particular scene parameter (S) is present, given some image attribute (I), and is obtained from Bayes's rule:

$$p(S|I) \propto p(I|S)p(S). \quad \text{(A5)}$$

In our specific example, the scene parameter we are estimating is slant (φ), and the image attributes are the orientation of the line (γ) and the disparity (d). Assuming independence between the perspective and the disparity cues, expression (A5) becomes

$$p(\varphi|\gamma, d; r, \sigma_d, \sigma_p) \propto p(\gamma|\varphi; r)p(d|\varphi; \sigma_d)p(\varphi; \sigma_p). \quad \text{(A6)}$$

E.   Origin of the Two Modes
Our model works by switching between strong and weak rectangularity assumptions. This is achieved by exchanging the spread of the line orientation on the surface (θ) in Eq. (A1) between a parameter r_1 and a parameter r_2. Therefore two posterior distributions will be obtained, following these two perspective likelihoods. The disparity likelihood and the prior stay the same for the two modes.

F.   Combination of the Two Modes
A decision is reached by combining the two posterior distributions and subjecting this combination to a gain function that makes the system robust to noise (see text for details). The gain function is a Gaussian centered on zero with the spread (σ_g) left as a free parameter:

$$G(\varphi; \sigma_g) = \frac{1}{\sqrt{2\pi\sigma_g{}^2}}\exp\left(\frac{-\varphi^2}{2\sigma_g{}^2}\right). \quad \text{(A7)}$$

This gain function is convolved with the sum of the two posteriors, to give the expected gain $E(\varphi)$:

$$E(\varphi; r_1, r_2, \sigma_d, \sigma_p, \sigma_g)$$
$$= G(\varphi; \sigma_g) * [p(\varphi|\gamma, d; r_1, \sigma_d, \sigma_p)$$
$$+ p(\varphi|\gamma, d; r_2, \sigma_d, \sigma_p)]. \quad \text{(A8)}$$

The model's predictions are then simply the locations of the peaks of this expected gain function. A single peak

corresponds to observers making a single response. If there are two peaks, the model predicts two separate responses corresponding to disparity- and perspective-dominated percepts.

REFERENCES AND NOTES

1. These assumptions are often unnoticed, and the prior knowledge is not something the observer needs to be aware of.[2] Bayesian theory provides a general framework that incorporates such assumptions.
2. H. von Helmholtz, *Handbuch der Physiologischen Optik* (Voss, Hamburg, Germany, 1866), Vol. 3, Sec. 26.
3. R. S. Allison and I. P. Howard, "Temporal dependencies in resolving monocular and binocular cue conflict in slant perception," Vision Res. **40**, 1869–1886 (2000).
4. R. S. Allison and I. P. Howard, "Stereopsis with persisting and dynamic textures," Vision Res. **40**, 3823–3827 (2000).
5. B. J. Gillam, "Perception of slant when perspective and stereopsis conflict: experiments with aniseikonic lenses," J. Exp. Psychol. **78**, 299–305 (1968).
6. B. J. Gillam and C. Ryan, "Perspective, orientation disparity, and anisotropy in stereoscopic slant perception," Perception **21**, 427–439 (1992).
7. C. Ryan and B. Gillam, "Cue conflict and stereoscopic surface slant about horizontal and vertical axes," Perception **23**, 645–658 (1994).
8. B. J. Gillam and M. L. Cook, "Perspective based on stereopsis and occlusion," Psychol. Sci. **12**, 424–429 (2001).
9. A. H. Smith, "Perceived slant as a function of stimulus contour and vertical dimension," Percept. Mot. Skills **24**, 167–173 (1967).
10. R. van Ee, M. S. Banks, and B. T. Backus, "An analysis of binocular slant contrast," Perception **28**, 1121–1145 (1999).
11. M. S. Banks and B. T. Backus, "Extra-retinal and perspective cues cause the small range of the induced effect," Vision Res. **38**, 187–194 (1998).
12. W. M. Youngs, "The influence of perspective and disparity cues on the perception of slant," Vision Res. **16**, 79–82 (1976).
13. C. Wheatstone, "Contributions to the physiology of vision—part the first. On some remarkable and hitherto unobserved phenomena of binocular vision," Philos. Trans. R. Soc. London **128**, 371–394 (1838).
14. W. Schriever, "Experimentelle Studien über stereoskopisches Sehen," Z. Psychol. Physiol. Sinnesorgane **96**, 113–170 (1925).
15. K. A. Stevens, M. Lees, and A. Brookes, "Combining binocular and monocular curvature features," Perception **20**, 425–440 (1991).
16. H. Hill and V. Bruce, "Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion," Perception **22**, 887–897 (1993).

17. R. van Ee, K. Hol, and C. J. Erkelens, "Bistable stereoscopic percepts and depth cue combination," Perception **30**, S42 (2001).
18. T. V. Papathomas, "Experiments on the role of painted cues in Hughes's reverspectives," Perception **31**, 521–530 (2002).
19. See also other interesting contributions in Refs. 20–23.
20. R. Gregory, *The Intelligent Eye* (Weidenfeld and Nicholson, London, 1970).
21. J. Slyce, *Patrick Hughes: Perverspective* (Momentum, London, 1998).
22. N. J. Wade and P. Hughes, "Fooling the eyes: *trompe l'oeil* and reverse perspective," Perception **28**, 1115–1119 (1999).
23. E. Mach, "Über die physiologische Wirkung räumlich verteilter Lichtreize," Sitzungsber. d. Wiener Akad. **54**, 3 (1866).
24. R. van Ee, L. C. J. van Dam, and C. J. Erkelens, "Bistability in perceived slant when binocular disparity and monocular perspective specify different slants," J. Vision **2**, 597–607 (2002).
25. Although eye movements play a role, the perceptual bistability seems to be predominantly central. We are currently measuring eye movements while subjects experience bistability in our grid stimuli. Our preliminary findings reveal that switching between the two percepts can occur by effort of will while subjects keep strict fixation. When eye movements are allowed, there is no clear correlation between perceptual flips and both eye movements and blinks.[26]
26. L. C. J. van Dam and R. van Ee, "Bistability in stereoscopically perceived slant about a horizontal axis," J. Vision (to be published) (Abstract Book VSS03).
27. R. van Ee and W. Richards, "A planar and a volumetric test for stereoanomaly," Perception **31**, 51–64 (2002).
28. R. van Ee and C. J. Erkelens, "Temporal aspects of binocular slant perception," Vision Res. **36**, 43–51 (1996).
29. A sensible objection to this metrical slant-estimation method is that it is hard to interpret the data because a slant angle that is estimated at 35 deg in one trial might look like 40 deg in another trial. Previous work has demonstrated, however, that subjects have a relatively constant internal reference and that they do not regard this task as difficult. This estimation method has been used previously for real planes[10] and when subjects wore distorting lenses.[30] In addition, a similar metrical depth-estimation method was successfully used for volumetric stimuli.[31]
30. W. J. Adams, M. S. Banks, and R. van Ee, "Adaptation to three-dimensional distortions in human vision," Nat. Neurosci. **4**, 1063–1064 (2001).
31. R. van Ee and B. L. Anderson, "Motion direction, speed, and orientation in binocular matching," Nature **410**, 690–694 (2001).
32. Bayesian theory is a rich mathematical theory.[33,34] Massaro[35] and Clark and Yuille[36] made Bayesian theory accessible to speech perception and visual perception, respectively. See also excellent chapters in Refs. 37 and 38 and introductory tutorials in Refs. 39 and 40 on applications in visual cue integration.
33. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, Berlin, 1985).
34. T. Ferguson, *Mathematical Statistics: a Decision Theoretic Approach* (Academic, New York, 1967).
35. D. W. Massaro, *Speech Perception by Ear and Eye* (Erlbaum, Hillsdale, N.J., 1987).
36. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems* (Kluwer Academic, Boston, 1990).
37. L. T. Maloney, "Statistical decision theory and biological vision," in *Perception and the Physical World*, D. Heyer and R. Mausfeld, eds. (Wiley, Chichester, UK, 2002).
38. A. L. Yuille and H. H. Bülthoff, "Bayesian decision theory and psychophysics," in *Perception as Bayesian Inference*, D. C. Knill and W. Richards, eds. (Cambridge U. Press, Cambridge, UK, 1996).
39. D. C. Knill, D. Kersten, and A. L. Yuille, "Introduction: a Bayesian formulation of visual perception," in *Perception as Bayesian Inference*, D. C. Knill and W. Richards, eds. (Cambridge U. Press, Cambridge, UK, 1996).
40. P. Mamassian, M. S. Landy, and L. T. Maloney, "Bayesian modelling of visual perception," in *Probabilistic Models of the Brain*, R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, eds. (MIT, Cambridge, Mass., 2002).
41. D. C. Knill and W. Richards, *Perception as Bayesian Inference* (Cambridge U. Press, Cambridge, UK, 1996).
42. J. Porrill, J. P. Frisby, W. J. Adams, and D. Buckley, "Robust and optimal use of information in stereo vision," Nature **397**, 63–66 (1999).
43. H. H. Bülthoff and H. A. Mallot, "Integration of stereo, shading and texture," in *AI and the Eye*, A Blake and T. Troscianko, eds. (Wiley, New York, 1990).
44. W. T. Freeman, "The generic viewpoint assumption in a framework for visual perception," Nature **368**, 542–545 (1994).
45. W. T. Freeman, "The generic viewpoint assumption in a Bayesian framework," in *Perception as Bayesian Inference*, D. C. Knill and W. Richards, eds. (Cambridge U. Press, Cambridge, UK, 1996).
46. H. H. Bülthoff and A. L. Yuille, "Shape from X: psychophysics and computation," in *Sensor Fusion III: 3D Perception and Recognition*, P. S. Schenker, ed., Proc. SPIE **1383**, 235–246 (1990).
47. H. H. Bülthoff, "Shape from X: psychophysics and computation," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT, Cambridge, Mass., 1991).
48. A. L. Yuille, D. Geiger, and H. H. Bülthoff, "Stereo integration, mean field theory and psychophysics," Network **2**, 423–442 (1991).
49. D. Ascher and N. M. Grzywacz, "A Bayesian model for the measurement of visual velocity," Vision Res. **40**, 3427–3434 (2000).
50. M. A. Hogervorst and R. A. Eagle, "Biases in three-dimensional structure-from-motion arise from noise in the early visual system," Proc. R. Soc. London Ser. B **265**, 1587–1593 (1998).
51. L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," Vision Res. **39**, 2729–2737 (1999).
52. P. Mamassian and M. S. Landy, "Observer biases in the 3D interpretation of line drawings," Vision Res. **38**, 2817–2832 (1998).
53. P. Mamassian and M. S. Landy, "Interaction of visual prior constraints," Vision Res. **41**, 2653–2668 (2001).
54. J. C. A. Read, "A Bayesian model of stereopsis depth and motion direction discrimination," Biol. Cybern. **86**, 117–136 (2002).
55. It is of historical interest to note that Bayes died in 1761 and that an essay that Bayes wrote had been published by the Royal Society[56] two years after his death. Bayes's theorem was originally developed to model human conscious judgments during the playing of games, but it has proven to be wrong for this purpose.[37] In modern vision science, Bayes's work has been attached to the following equation: $p(S|I) \propto p(I|S)p(S)$. It therefore comes as a surprise that this equation is not present in Bayes's essay. According to Dale,[57] Laplace's[58] formulations have mistakenly been applied as those of Bayes. This is not to say that Bayes does not deserve the name for the theory.
56. T. Bayes, "An essay towards solving a problem in the doctrine of chances," Philos. Trans. R. Soc. London **53**, 370–418 (1763).
57. A. I. Dale, "Bayes or Laplace? An examination of the origin and early applications of Bayes' theorem," Arch. Hist. Exact Sci. **27**, 23–47 (1982).
58. P. S. Laplace, *Théorie Analytique des Probabilités* (Courcier, Paris, 1812).
59. To decide which of the peaks corresponds to the weak rectangularity mode and which of the peaks corresponds to the strong rectangularity mode, we compared the peaks in the expected gain distribution with the highest peaks in the individual posterior distributions. It is relatively straight-

forward to shift the bifurcation point by applying a different gain function, producing bifurcation points that perfectly fit the obtained data.   However, the coefficient goodness of fit that we generally applied (see Table 1) becomes then slightly worse relative to the best fit of the model.

60. D. Kersten, H. H. Bülthoff, B. L. Schwartz, and K. J. Kurtz, "Interaction between transparency and SFM," Neural Comput. **4**, 573–589 (1992).
61. In the top left panel of Fig. 3 the model prediction exceeds the disparity-specified slant.   This overprediction is relatively easy to prevent, but it involves, to our mind, *ad hoc* physiological assumptions.
62. M. L. Braunstein, G. J. Andersen, M. W. Rouse, and J. S. Tittle, "Recovering viewer-centered depth from disparity, occlusion, and velocity gradients," Percept. Psychophys. **40**, 216–224 (1986).
63. H. H. Bülthoff and H. A. Mallot, "Integration of depth modules: stereo and shading," J. Opt. Soc. Am. A **5**, 1749–1758 (1988).
64. B. A. Dosher, G. Sperling, and S. A. Wurst, "Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure," Vision Res. **26**, 973–990 (1986).
65. B. J. Rogers and T. S. Collett, "The appearance of surfaces specified by motion parallax and binocular disparity," Q. J. Exp. Psychol. A **41**, 697–717 (1989).
66. J. Turner, M. L. Braunstein, and G. J. Andersen, "Relationship between binocular disparity and motion parallax in surface detection," Percept. Psychophys. **59**, 370–380 (1997).
67. H. C. van der Meer, "Interrelation of the effects of binocular disparity and perspective cues on judgments of depth and height," Percept. Psychophys. **29**, 481–488 (1979).
68. C. Wheatstone, "The Bakerian lecture: contributions to the physiology of vision—part the second.   On some remarkable and hitherto unobserved phenomena of binocular vision," Philos. Trans. R. Soc. London **142**, 1–17 (1852).
69. R. B. Freeman, "Theory of cues and the psychophysics of visual space perception," Psychonom. Monogr. **3**, 171–181 (1970).
70. L. T. Maloney and M. S. Landy, "A statistical framework for robust fusion of depth information," in *Visual Communications and Image Processing IV*, W. A. Pearlman, ed., Proc. SPIE **1199**, 1154–1163 (1989).
71. M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: in defense of weak fusion," Vision Res. **35**, 389–412 (1995).
72. I. Fine and R. A. Jacobs, "Modeling the combination of motion, stereo, and vergence angle cues to visual depth," Neural Comput. **11**, 1297–1330 (1999).
73. T. Poggio, E. B. Gamble, and J. J. Little, "Parallel integration of vision modules," Science **242**, 436–440 (1988).
74. R. van Ee and C. J. Erkelens, "Conscious selection of bistable 3D percepts described by neural population codes," J. Vision **2**, S549a (2002).