CURRENT OPINION
www.current-opinion.com

Bayesian models of object perception

Daniel Kersten* and Alan Yuille†

The human visual system is the most complex pattern recognition device known. In ways that are yet to be fully understood, the visual cortex arrives at a simple and unambiguous interpretation of data from the retinal image that is useful for the decisions and actions of everyday life. Recent advances in Bayesian models of computer vision and in the measurement and modeling of natural image statistics are providing the tools to test and constrain theories of human object perception. In turn, these theories are having an impact on the interpretation of cortical function.

Addresses
*Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, USA
e-mail: kersten@umn.edu
†Departments of Statistics and Psychology, University of California, Los Angeles 8130 Math Sciences Building, UCLA, Box 951554, Los Angeles, CA 90095-1554, USA
e-mail: yuille@stat.ucla.edu

Current Opinion in Neurobiology 2003, **13**:150–158

This review comes from a themed issue on
Cognitive neuroscience
Edited by Brian Wandell and Anthony Movshon

0959-4388/03/$ – see front matter

DOI 10.1016/S0959-4388(03)00042-4

Abbreviations

fMRI	function magnetic resonance imaging
I	image description or feature
p	probability
r	reliability
S	object description
V1	primary visual cortex

Introduction

The certainty of human visual experience stands in stark contrast to the objective ambiguity of the features of natural images, that is, the images received by the eye during daily activities. Similar objects can give rise to very different images (Figure 1a), whereas different objects can give rise to practically identical images (Figure 1b). Although we can easily see the bicycles in (Figure 1a (top), the images themselves are very complex (see (Figure 1a [bottom]), and the features of these images, corresponding to the wheels and frames of the bicycles, are highly ambiguous. Our brains are specialized for understanding natural images, and this disguises the difficulty of dealing with their complexity and ambiguity.

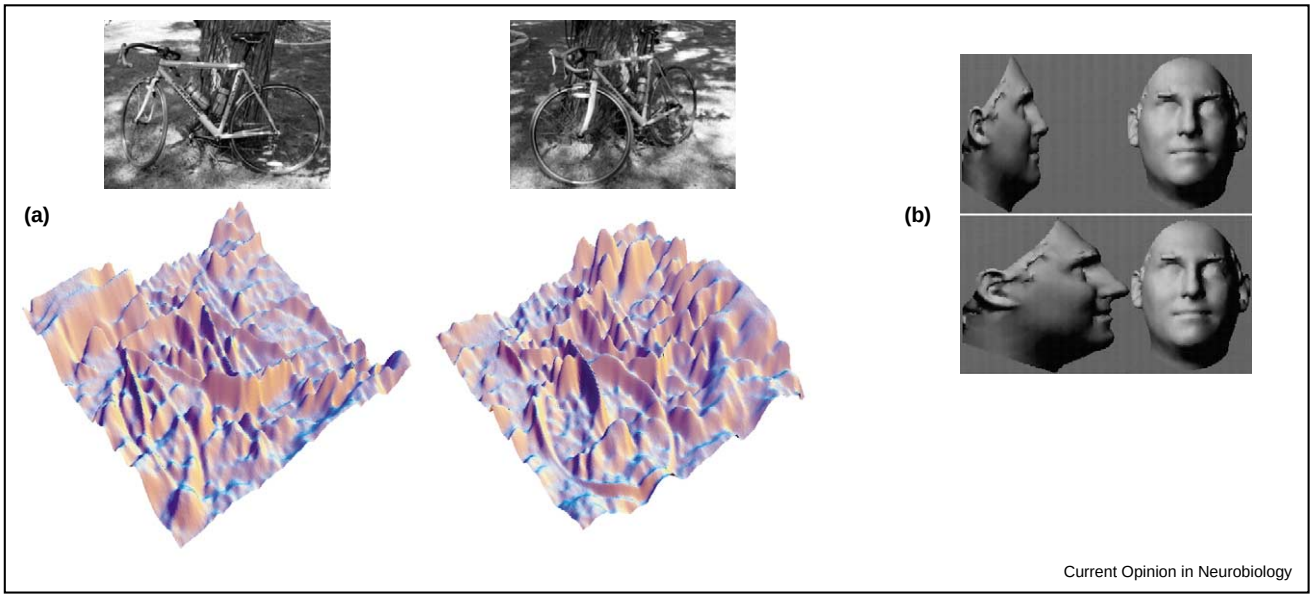
Visual scientists must come to terms with this ambiguity and complexity. Computer vision is a relatively recent science that develops theories and algorithms to extract information from natural images useful for recognition, scene interpretation, and robot actions. One of the recent lessons from computer vision is that natural images have properties and structures that differ greatly from the artificial stimuli typically studied by visual scientists (compare (Figure 1a [bottom]) to the dots, sinusoidal gratings, and line drawings traditionally used as experimental stimuli). Nevertheless, the neural and psychological study of visual perception requires us to simplify problems so that they can be investigated under controlled circumstances. Bayesian models of visual perception allow scientists to break these problems down into limited classes of categories that lie within a theoretical framework that can be extended to deal with the ambiguities and complexities of natural images in studies of computer vision.

The Bayesian framework for vision has its origins with Helmholtz's notion of unconscious inference [1], and in recent years it has been formally developed by visual scientists [2,3,4•,5••]. It uses Bayesian probability theory [6], in which prior knowledge about visual scenes is combined with image features to infer the most probable interpretation of the image (Figure 2). The Bayesian approach can be used to derive statistically optimal 'ideal observer' models, which can be used to normalize human performance with respect to the information needed to perform a visual task [7–9]. There is growing evidence, some of which is reviewed below, showing that human visual perception can be close to ideal for visual tasks of high utility and under visual conditions that approximate those typically encountered. The Bayesian approach to human object perception has been recently advanced along two main fronts: the analysis of real-world statistics, and a categorization and better understanding of inference problems. Bayesian inference of object properties relies on probabilistic descriptions of image features as a function of their causes in the world, such as object shape, material, and illumination. Bayesian inference in addition relies on 'prior' descriptions of these same causes independent of the images. Computer vision studies have shown how measurements of real-world statistics, both of images and causes in the world, provide the basis on which to model the probabilities required to make reliable inferences of object properties from image features.

Real-world statistics

The statistical regularities in natural images and scene properties are essential for taming the complexity and ambiguity of image interpretation. For example, in

Figure 1



Objective ambiguity of images **(a)** The top panel shows two views of the same bicycle and the bottom panel plots the two images as surfaces (Reproduced with permission from [65]). **(b)** With different lighting directions, flattened and elongated facial geometries (the two left profiles) produce identical images (right). Despite this bas-relief ambiguity, the visual system arrives at a subjectively unique perceived shape, which in this case is typically different to either of the profiles. The profile views of the faces on the left illustrate the true depths of the surfaces. The bottom surface is the same as the top except for a scale factor in depth. By illuminating the top face from a more vertical angle than the bottom face, one obtains similar images. When the surface reflectance is uniform Lambertian, the images are identical. This trade-off between depth scale and illumination direction is called the bas-relief ambiguity. (Adapted from [63], © 1993 IEEE.)

natural images, nearby pixels tend to have the same intensity. Further, the distribution of the difference in intensity between pairs of pixels is highly non-Gaussian (Figure 3). Scene properties such as object shape, material, and illumination also show statistical regularities. A common assumption in computer vision is that surface orientation of an object tends to vary smoothly. Recent studies of natural image data have shown how statistics can provide an efficient characterization of homogeneous textures [10,11•,12], help to discount the illuminant [13,14•], find object boundaries [15•], contribute to scene recognition [16], and constrain the context for object recognition [17•]. Statistics on the geometry of contours have been used to explain aspects of human perception [18•,19•], and statistical shape regularities enable computer vision systems to group image features consistent with the constraints identified by Gestalt psychologists, such as requiring that nearby edges with similar orientations are likely to belong to the same contour [20].

Other statistical regularities relate image features to measurements of object or scene properties. They constrain local boundary detection (Figure 3a; [21•]), identify regularities on the changes of image features as illumination conditions vary (Figure 3b; [22]), and help to interpret human body motion [23]. Relating image intensity to measured surface-depth statistics has yielded computer vision solutions for face recognition [24], provided objective models [25] for face recognition experiments [26], and given insight into the functional nature of certain kinds of illusions [27•]. It is largely an open question of how the human visual system learns the appropriate statistical priors, but some priors as well as strategies for learning priors are likely to be rooted in our genes [28,29].

Basic Bayes

The Bayes formula for inverse inference is:

$$p(S|I) = \frac{p(I|S)p(S)}{p(I)}$$

where p(I|S) is the model for forming an image (I), and p(S) is prior knowledge about the naturally occurring structures (S) (Figure 2). Both p(S) and p(I|S) can be learnt from real-world statistics.

The distributions of p(I|S) and p(S) define an ensemble of problem instances governed by a joint probability $p(S,I) = p(I|S)p(S)$. This probability can be represented by a simple graph, which we call Basic Bayes, in which the two nodes of the graph, represented by the two circles, represent S and I (Figure 4a). Ideal observers can be defined for this ensemble of problems. An ideal observer infers the most probable (or in the more general case of Bayesian decision theory, the most useful) value of S

Figure 2

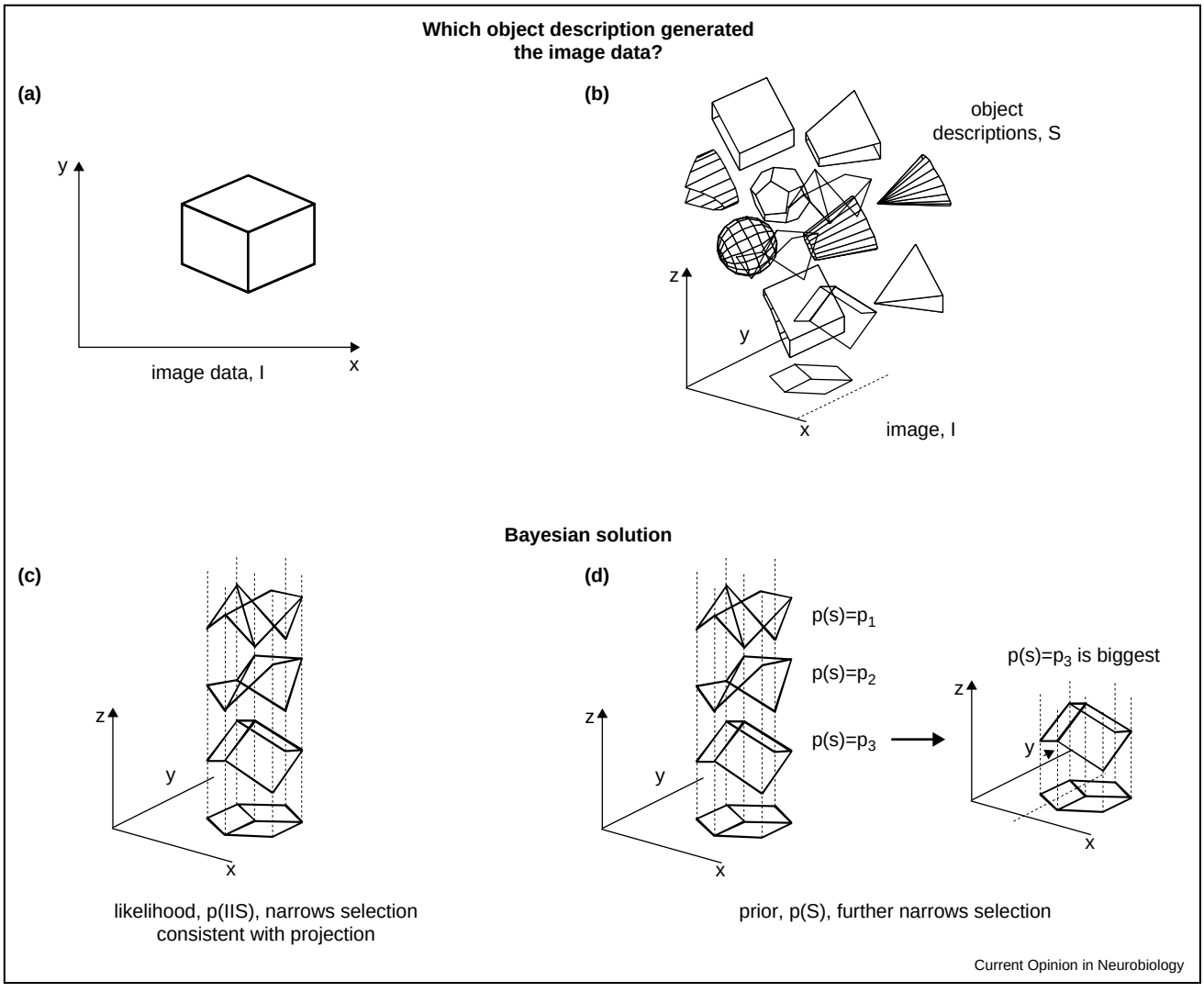


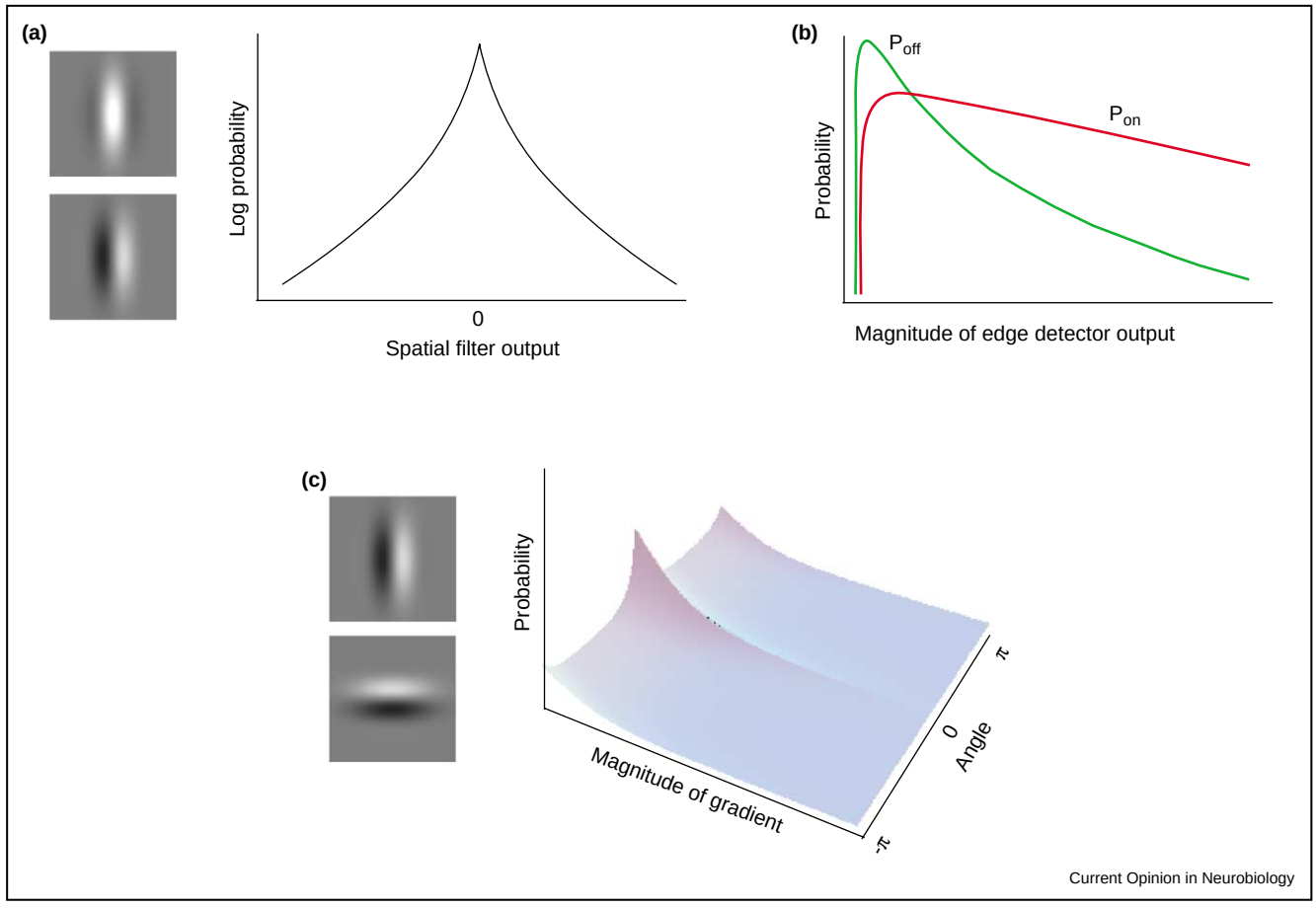
Illustration of Bayesian object perception **(a)** What 3D object caused the image of a cube? The likelihood (p(I|S)=p(image data|object descriptions)) constrains the possible set of objects to those consistent with the image data, but even this is an infinite set. **(b)** The prior knowledge probability (p(S)=p(object descriptions)) constrains the consistent set of 3D objects to those that are more probable in the world. **(c,d)** The probability over all instances is determined by the product of the likelihood and prior knowledge: p(S,I)=p(I|S)p(S)=p(object descriptions, image data). See the section on Basic Bayes for mathematical definitions. (Adapted from [64], © 1993 IEEE.)

given I, that is, the value of S that gives the maximum value of p(S|I). The ideal's performance is determined by the information for the visual task and can be used as a benchmark to evaluate human performance.

One consequence of Baye's formula is that perception is a trade-off between image feature reliability, as embodied by p(I|S), and the prior p(S). Some perceptions may be driven more by prior knowledge and some more by data. The less reliable the image features (e.g. the more they are ambiguous), the more the perception is influenced by the prior knowledge. This trade-off has been illustrated for a variety of visual phenomena [30,31,32••,33].

We can obtain a deeper understanding by categorizing Bayesian problems with graphs, or influence diagrams. First, decompose the natural world S into components S_1, S_2, S_3. . .., the image into features I_1, I_2, I_3,. . .., and express the ensemble distribution as $p(S, I) = p(S_1, S_2, \ldots, I_1, I_2, \ldots)$. Such a decomposition may not be straightforward and some ensemble distributions may be hard to express or to learn; however, for a given problem we can begin by characterizing how variables in the world influence each other and the resulting image features by representing the distribution with an influence diagram in which the nodes correspond to the variables S_1, S_2, S_3,. . .., and I_1, I_2, I_3, ..., and we draw links between the nodes that directly

Figure 3



Examples of natural image statistics **(a)** The far-left upper panel illustrates models of receptive fields of neurons in V1 that can be used to compute image features. The center panel shows the typical histogram, or probability distribution, of the feature response to a natural image. This standard form, where zero feature response is most common and larger responses occur with exponential fall-off, has been reported for many visual features and used to motivate neural encodings hypotheses [11•]. To perform Bayesian inference, it is helpful to characterize feature responses according to the world events that give rise to them. The sum of the squares of these filter responses can be used as an edge detector filter. **(b)** The histograms of an edge detector filter, P_{on} and P_{off}, conditioned on whether the filter is evaluated on or off an edge (e.g. an object boundary) in the scene. We can then infer the positions of edges in an image by calculating edge-filter responses and evaluating the ratio P_{on}/P_{off} — the higher the ratio the more likely there is an edge [21•]. **(c)** This determines directional image features that are statistically insensitive to changes in illumination and can be used to design object recognition systems that are roughly insensitive to illumination [22]. A directional filter is specified by two filters whose receptive fields are tuned at 90° to each other (bottom left). Its response is a two-dimensional vector that can be characterized by its magnitude and angle. The plot shows the probability of the changes in magnitude and angle as the illumination conditions vary. It shows that the angle direction is very insensitive to the illumination whereas the magnitude is more variable. Hence, the angle can be treated as a statistical invariant and used as input to a recognition system.

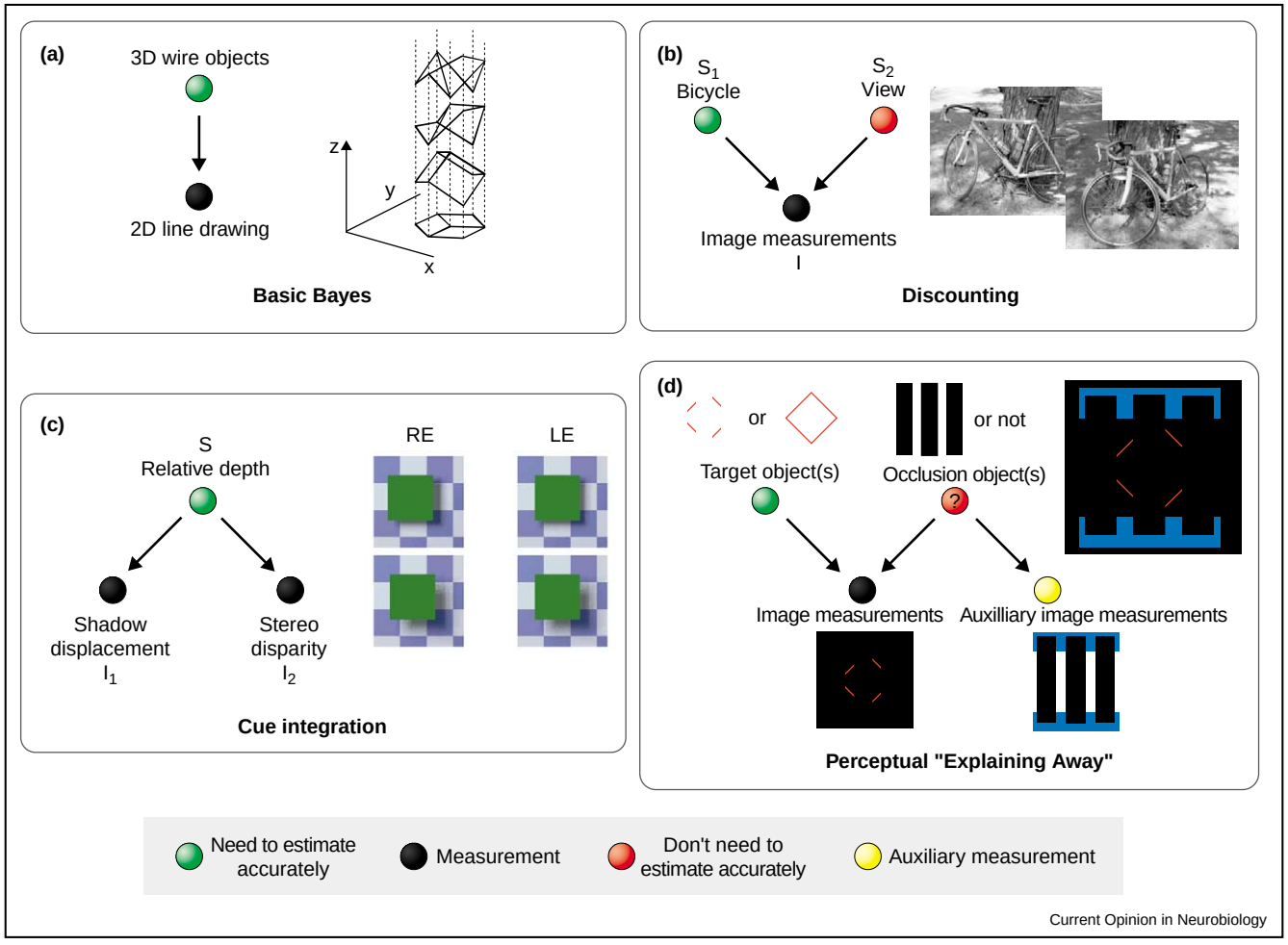
influence each other (in the worst case, all nodes are connected and we get a complicated diagram). These influence diagrams provide a classification of visual inference problems in terms of how image features are generated (Figure 4). We now describe three important categories of visual inference problems.

Discounting

How does the visual system enable us to infer the same object despite considerable image variation caused by confounding variables, such as viewpoint, illumination, occlusion, and background changes? This is the well-known problem of invariance (Figure 1a). Confounding variables are analogous to 'noise' in classical signal-detection theory, but they are more complicated to model and they affect image formation in a non-linear manner.

From a Bayesian perspective, we can model problems with confounding variables by the influence diagram in (Figure 4b. We define an ensemble distribution $p(S_1, S_2, I_1)$, where S_1 is the target (e.g. a bicycle), S_2 is the

Figure 4



Influence diagrams representing generative models for four categories of Bayesian inference. The arrows indicate how scene or object properties (S) influence image measurements (I). Visual inference goes against the arrows (i.e. it is inverse inference). The influence diagrams simplify the problem structure, determining how the joint probability over problem instances can be factored. Visual inference also depends on the task, and the node colors represent random variables that fall into four classes. The variables may be: 1) known (black); 2) unknown and need to be estimated accurately (green); 3) unknown, but do not need to be accurately estimated (red); 4) auxiliary, not directly influenced by the object variable of interest, but may be useful for resolving ambiguity (yellow). **(a)** Basic Bayes illustrates how a scene variable influences an image measurement. The influence diagram factoring is $p(S,I)=p(I|S)p(S)$. The cube problem in Figure 3 is an example. **(b)** Discounting illustrates two scene factors (S_1 and S_2) that both influence the image measurement (I). The one that does not need to be estimated is a confounding variable to be discounted. The influence diagram structure corresponds to $p(S_1, S_2,I)=p(I|S_1, S_2)p(S_1)p(S_2)$. Recognizing that the images are of the same bicycle (Figure 1a) requires one to discount a change in view. **(c)** Cue integration shows how the same factor (S) in a scene influences two different features or cues (I_1 and I_2). The influence diagram structure corresponds to $p(S, I_1, I_2)=p(I_1, I_2|S)p(S)$. The shadow means that the lower two green squares appear to be further from the checkerboard; however, when seen in stereo (with eyes crossed), the disparity and shadow cues combine, and the upper green square is seen to be further from the checkerboard. **(d)** Perceptual 'explaining away'. Two scene parameters (S_1 and S_2) may influence an image measurement (I_1), and an auxiliary measurement (I_2) 'tips the balance' in favor of a different value of S_1. The influence diagram structure corresponds to $p(S_1, S_2, I_1, I_2)=p(I_2|S_2)p(I_1|S_1, S_2)p(S_1)p(S_2)$. Four red line segments may appear as four distinct objects because the vertices are occluded. When auxiliary evidence (the blue bars) is taken into account, the missing vertices are explained and the four red-line segments become perceptually organized into a single diamond [50,52••].

confounding variable (e.g. viewpoint), and I_1 is the image. Then we discount the confounding variables by integrating them out (or summing over them):

$$p(S_1, I_1) = \sum_{S_2} p(S_1, S_2, I_1)$$

Discounting illumination by integrating it out can reduce ambiguity with regards to the location or shape of an object [34,35]. Integrating out illumination level or direction has also been used to model apparent surface color [36,37]. Similarly, viewpoint can be treated as a confounding variable that can be discounted [8].

From the perspective of utility theory, discounting a variable is equivalent to treating it as having such low utility that it is not worth estimating [38]. At the other extreme, an alternative to 'integrating out' across a variety of possibilities is to accurately estimate the confounding variable that accounts for this particular image. This can result in the explaining away phenomenon discussed below. The choice of which variables to discount will depend on the task. Illumination is a confounding variable if the task is to recognize or to determine the depth of an object, but not if the task is to determine the direction of the light source. This task dependence may account for studies in which different 'operationalisations' for measuring a human's estimates of shape lead to inconsistent, although related, results [39].

Cue integration

Vision integrates information from a variety of sources about the relative depths of objects, their shapes and their motions. For example, one can identify more than a dozen cues that the human visual system utilizes for depth perception. These can be modeled from a Bayesian perspective [40].

We illustrate cue integration in the influence diagram in (Figure 4c. An important situation arises when the nodes represent gaussian variables and we have estimates $\hat{S}_1$ and $\hat{S}_2$ for each cue alone (i.e. $\hat{S}_1$ is the best estimator for $p(S|I_1)$). Then, optimal integration (i.e. the most probable value) of the two estimates takes into account the uncertainty caused by measurement noise (the standard deviations, σ_1 and σ_2), and is given by the weighted average,

$$\hat{S}_{optimal} = \hat{S}_1 \frac{r_1}{r_1 + r_2} + \hat{S}_2 \frac{r_2}{r_1 + r_2}$$

where r_i, the reliability, is the reciprocal of the variance ($r_i = 1/\sigma_i^2$). This model has been used to study whether the human visual system combines cues optimally [41•]. In particular, visual and haptic information about object size are combined, and weighted according to the reliability of the source [42••]. A more complicated model uses robust statistics to determine whether one measurement is an outlier, and therefore should not be integrated with the other measurement [43]. Integration is also important for grouping local image elements that are likely to belong to the same surface. A model of Bayes optimal motion integration accounts for a large number of motion illusions and experimental data [32••,44]. In another domain, the human visual system optimally combines spatial frequency and orientation information when detecting the boundary between two regions [45]. Human observers behave like an optimal observer when integrating information from skew symmetry and disparity in perceiving the orientation of a planar object [46•]. Psychological experiments show that prior probabilities can also combine like weighted cues [47•].

However, not all kinds of cue integration are consistent with the simple influence diagram of (Figure 4c. It has been argued [38] that strong coupling of visual cues [40] is required to model a range of visual phenomena [48].

Perceptual 'explaining away'

Visual ambiguity can be reduced by auxiliary measurements that may be available in a given image or actively sought (see the influence-diagram structure in (Figure 4d). Indeed, these measurements may completely alter the percept by explaining away the effects of the confounding variables. The term 'explaining away' originates in the context of reasoning where a change in the probability of one competing hypothesis affects the probability of another [49].

A striking example of perceptual 'explaining away' can be experienced when the diamond in (Figure 4d translates back and forth horizontally; yet depending on auxiliary evidence, the segments may appear to be moving vertically [50,51•,52••]. In other examples, binocular stereo information (effectively adding a new image of the object) can be used to change the apparent shape of a card folded concavely and, as a consequence, can also change the apparent color saturation [37]. However, human perception can also unexpectedly fail to explain away [53].

Explaining away is closely related to the situation seen with competitive models, in which two alternative models attempt to explain the same data. It has been argued that this accounts for a range of visual phenomena [38], including the estimation of material properties [54]. This approach has been used successfully in computer vision systems [55••].

Neural implications

What are the neural implications of ideal-observer models? They can, of course, be treated as purely functional models of perception. They will be far more plausible, however, if they can be linked to neural mechanisms and used to make experimental predictions. Fortunately, the graphical structure of these models often makes it straightforward to map them onto networks and suggests neural implementations.

One class of Bayesian models can be implemented by parallel networks with local interactions. These include the motion models [32••] and a temporal-motion model [56], which was designed to be consistent with neural mechanisms. In these models, the prior knowledge and likelihood functions are implemented by synaptic weights. Such models are broadly consistent with some electrode recordings [57••], but detailed testing is impractical at present.

There are proposed neural mechanisms for representing uncertainty in neural populations that thereby give a

mechanism for weighted cue combination. The most plausible candidate is population encoding [58].

Interestingly, the graphical nature of Bayesian models is suggestive of the feedforward and feedback connections [59,60•] that connect the visual areas of primates. A possible role for higher-level visual areas, for example the lateral occipital complex [61], may be to represent hypotheses regarding object properties that could be used to resolve ambiguities in the incoming retinal-image measurements that are represented in the primary visual cortex (V1). These hypotheses could predict incoming data through feedback and be tested by computing a difference signal or residual at the earlier level [59]. Thus, low activity at an early level would mean a 'good fit' or explanation of the image measurements. Experimental support for this possibility comes from recent functional magnetic resonance imaging (fMRI) data [52••].

Two alternative theoretical possibilities might explain why early visual activity is reduced. High-level areas may explain away the image and cause the early areas to be completely suppressed; that is, high-level areas tell lower levels to 'shut up'. Alternatively, high-level areas might sharpen the responses of the early areas by reducing activity that is inconsistent with the high level interpretation; that is, high-level areas tell lower levels to 'stop gossiping'. This second possibility seems more consistent with some electrode-recording experiments [57••], but further experiments are needed to resolve this issue.

Moreover, there are alternative ways to implement Bayesian models even if feedback and feedforward connections are used. A recent approach uses image features to predict high-level structure that can then be verified by comparison to the image, and has been very successful for segmenting natural images [62•].

Conclusions

The Bayesian approach has proven very useful for both designing ideal observer models for psychological experiments and designing practical computer-vision systems. By taking advantage of the statistical regularities in images and scene structures this approach offers a way to deal with the ambiguity and complexity of natural images. Different visual tasks can be characterized by graphical models that illustrate the dependencies between the variables that describe the task. These graphical models also suggest ways to implement these functional models in terms of neural mechanisms that can be tested by fMRI and electrode recording.

Acknowledgements

Supported by grants from the National Institute of Health RO1 EY11507-001, EY02857 and EY12691, and National Science Foundation SBR-9631682.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Helmholtz H: *Handbuch der physiologischen optik*. Leipzig: Voss L; 1867. [Translation of title: Handbook of Physiological Optics.]

2. Grenander U: *Elements of Pattern Theory*. Baltimore: John Hopkins University Press; 1996.

3. Knill DC, Richards W: *Perception as Bayesian Inference*. Edited by Knill DC, Richards W. Cambridge: Cambridge University Press; 1996.

4. • Mamassian P, Landy MS, Maloney LT: **Bayesian modelling of visual perception** In *Probabilistic Models of the Brain: Perception and Neural Function*. Edited by Rao RPN, Olshausen BA, Lewicki MS. Cambridge, MA: MIT Press; 2002: 13-36.
This chapter, and [5••] have good introductory overviews of the Bayesian approach to vision.

5. •• Rao RPN, Olshausen BA, Lewicki MS: *Probabilistic Models of the Brain: Perception and Neural Function*. Edited by Rajesh PN, Rao R, Bruno A, Olshausen B, Michael S, Lewicki M. Cambridge, Mass: MIT Press; 2002.
This book gives a recent survey of current approaches to modeling and understanding brain function from a probabilistic perspective. It includes computationally motivated models, such as those described in this review, neuronally motivated models, and models of predictive coding using cortical feedback.

6. Bayes T: **Essay towards solving a problem in the doctrine of chances**. *Philosop Trans Roy Soc* 1763, **53**:370-418.

7. Barlow HB: **A method of determining the overall quantum efficiency of visual discriminations**. *J Physiol (Lond)* 1962, **160**:155-168.

8. Liu Z, Knill DC, Kersten D: **Object classification for human and ideal observers**. *Vision Res* 1995, **35**:549-568.

9. Eckstein MP, Thomas JP, Palmer J, Shimozaki SS: **A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays**. *Percept Psychophys* 2000, **62**:425-451.

10. Portilla J, Simoncelli EP: **A parametric texture model based on joint statistics of complex wavelet coefficients**. *Intl J Comput Vision* 2000, **40**:9-71.

11. • Simoncelli EP, Olshausen BA: **Natural image statistics and neural representation**. *Annu Rev Neurosci* 2001, **24**:1193-1216.
This is an excellent review of generic natural image statistics with a primary emphasis on early neural coding. Although the basic theme is related to the theme of this review, it does not focus on statistics for inference of specific object properties.

12. Zhu SC, Wu Y, Mumford D: **Minimax entropy principle and its applications to texture modeling**. *Neural Comput* 1997, **9**:1627-1660.

13. Judd DB, MacAdam DL, Wyszecki G: **Spectral distribution of typical daylight as a function of correlated color temperature**. *J Opt Soc Am* 1964, **54**:1031-1040.

14. • Golz J, MacLeod DI: **Influence of scene statistics on colour constancy**. *Nature* 2002, **415**:637-640.
We see objects as having constant surface color despite the fact that the retinal input is a function of both surface color and illumination. A reddish scene under white light can give rise to the same average stimulation as a neutral scene under red light. Using a spectral database of natural images, the authors showed that the correlation between redness and luminance could be used to disambiguate the scene from the lighting color, and further, that the weight given to the estimate of the illuminant was statistically optimal.

15. • Fine I, MacLeod DIA: **Visual segmentation based on the luminance and chromaticity statistics of natural scenes [abstract]**. *J Vision* 2001, **1(3)**:63a. URL: http://journalofvision.org/1/3/63, DOI 10.1167/1.3.63.
Differences in luminance between two points in a natural scene are statistically associated with differences in chromaticity. The authors incorporated this information into a Bayesian model to distinguish

whether two points are on the same surface or not, and showed that it predicted human performance. (This research will appear as an article in the *Journal of the Optical Society of America A*, Special Issue on Bayesian and Statistical Approaches to Vision to appear in 2003.)

16. Oliva A, Schyns PG: **Diagnostic colors mediate scene recognition**. *Cognit Psychol* 2000, **41**:176-210.

17. • Torralba A, Sinha P: **Statistical context priming for object detection**. In *Proceedings of the International Conference on Computer Vision, ICCVO1; Vancouver, Canada*: IEEE Computer Society; 2001:763-770.
Holistic statistics of images are used to determine image context and indicate where a particular object is likely to be found in an image. This models human experiments that show that subject performance is significantly better when the object is immersed in its typical context than when it is isolated.

18. • Geisler WS, Perry JS, Super BJ, Gallogly DP: **Edge co-occurrence in natural images predicts contour grouping performance**. *Vision Res* 2001, **41**:711-724.
The authors used spatial filtering to extract local edge fragments from natural images. They measured statistics on the distance between the element centers, the orientation difference between the elements, and the direction of the second element relative to the orientation of first (reference) element, and showed that a model derived from the measured statistics could quantitatively predict human contour detection performance.

19. • Elder JH, Goldberg RM: **Ecological statistics of gestalt laws for the perceptual organization of contours**. *J Vision* 2002, **2**:324-353. URL: http://journalofvision.org/2/4/5/, DOI 10.1167/2.4.5.
Statistics of contours are measured from hand-segmented images and used to put probability distributions on the proximity, continuation, and luminance of contour elements. The relative effectiveness of these cues for grouping contour segments into extended contours is evaluated.

20. Zhu SC: **Embedding gestalt laws in Markov random fields**. *IEEE T Pattern Anal* 1999, **21**:1170-1187.

21. • Konishi SM, Yuille AL, Coughlan JM, Zhu SC: **Statistical edge detection: learning and evaluating edge cues**. *IEEE Pattern Anal* 2003, **25**:57-74.
This paper measures the statistical effectiveness of different visual features for detecting edges and determines the optimal way to combine them. The statistics of feature responses are measured using image databases where human observers have determined the ground truth positions of the edges. [Also see: Konishi SM, Yuille AL, Coughlan JM, Zhu SC: *Fundamental bounds on edge detection: an information theoretic evaluation of different edge cues.* In *Proceedings Computer Vision and Pattern Recognition* CVPR'99; Fort Collins, Colorado: 1999.]

22. Chen HF, Belhumeur PN, Jacobs DW: **In search of illumination invariants**. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2000, **1**:254–261.

23. Troje NF: **Decomposing biological motion: a framework for analysis and synthesis of human gait patterns**. *J Vision* 2002, **2**:371-387.

24. Atick JJ, Griffin PA, Redlich AN: **Statistical approach to shape from shading: reconstruction of three-dimensional face surfaces from single two-dimensional images**. *Neural Comput* 1996, **8**:1321-1340.

25. Vetter T, Troje NF: **Separation of texture and shape in images of faces for image coding and synthesis**. *J Opt Soc Am A* 1997, **14**:2152-2161.

26. Leopold DA, O'Toole AJ, Vetter T, Blanz V: **Prototype-referenced shape encoding revealed by high-level aftereffects**. *Nat Neurosci* 2001, **4**:89-94.

27. • Howe CQ, Purves D: **Range image statistics can explain the anomalous perception of length**. *Proc Natl Acad Sci USA* 2002, **99**:13184-13188.
Why does a vertical line appear longer than the same line drawn horizontally? Although there have been several recent studies of the statistics of image intensities and contours, this paper takes the next step to directly compare image statistics on line length with measurements of the causes in the originating 3D scene. The authors found that the average ratio of the 3D physical length to the projected image length from real-world measurements shows the same pattern as perceived length. This result is in the spirit of the 1930s and '40s work of Egon Brunswik.

28. Fiser J, Aslin RN: **Statistical learning of new visual feature combinations by infants**. *Proc Natl Acad Sci USA* 2002, **99**:15822-15826.

29. Geisler WS, Diehl RL: **Bayesian natural selection and the evolution of perceptual systems**. *Philos Trans R Soc Lond B Biol Sci* 2002, **357**:419-448.

30. Bülthoff HH, Yuille A: **Bayesian models for seeing surfaces and depth**. *Comment Theor Biol* 1991, **2**:283-314.

31. Mamassian P, Goutcher R: **Prior knowledge on the illumination position**. *Cognition* 2001, **81**:B1-9.

32. •• Weiss Y, Simoncelli EP, Adelson EH: **Motion illusions as optimal percepts**. *Nat Neurosci* 2002, **5**:598-604.
This paper shows that even an ideal observer 'sees' illusions. The 'aperture problem' is a classic problem in motion perception — how do you combine locally ambiguous motion measurements into a single velocity for an object? The authors devise an ideal observer whose prior knowledge is that motions tend to be slow, and integrates local measurements according to their reliabilities. With this model (using a single free parameter per experiment), the authors account for a wide range of motion phenomena and results in human perception that are usually thought to reflect different neural mechanisms.

33. Feldman J: **Bayesian contour integration**. *Percept Psychophys* 2001, **63**:1171-1182.

34. Kersten D: **High-level vision as statistical inference**. In *The New Cognitive Neurosciences 2nd Edition*. Edited by Gazzaniga MS. Cambridge: MIT Press; 1999:353-363.

35. Freeman WT: **The generic viewpoint assumption in a framework for visual perception**. *Nature* 1994, **368**:542-545.

36. Brainard DH, Freeman WT: **Bayesian color constancy**. *J Opt Soc Am A* 1997, **14**:1393-1411.

37. Bloj MG, Kersten D, Hurlbert AC: **Perception of three-dimensional shape influences colour perception through mutual illumination**. *Nature* 1999, **402**:877-879.

38. Yuille AL, Bülthoff HH: **Bayesian decision theory and psychophysics**. In *Perception as Bayesian Inference*. Edited by Knill DC, Richards W. Cambridge: Cambridge University Press; 1996:123-161.

39. Koenderink JJ, van Doorn AJ, Kappers AM, Todd JT: **Ambiguity and the 'mental eye' in pictorial relief**. *Perception* 2001, **30**:431-448.

40. Clark JJ, Yuille AL: *Data Fusion for Sensory Information Processing*. Boston: Kluwer Academic Publishers; 1990.

41. • Jacobs RA: **What determines visual cue reliability?** *Trends Cogn Sci* 2002, **6**:345-350.
Combining information according to reliability raises the obvious question of the title, where does the information come from? This paper provides a nice overview of recent work on optimal cue combination in human vision together with a discussion of what determines the weighting for cue integration. The article also discusses integration in the context of Kalman filtering, which is a special case of Bayesian estimation.

42. •• Ernst MO, Banks MS: **Humans integrate visual and haptic information in a statistically optimal fashion**. *Nature* 2002, **415**:429-433.
Object size can be perceived both visually and by touch. When the information from vision and touch disagree, vision usually dominates. The traditional interpretation that this is because 'vision is the dominant sense' is not particularly informative. The authors showed that when one takes into account the reliability of the sensory measurements, information from vision and touch are integrated optimally. Visual dominance occurs when the reliability of the visual estimation is greater than that of the haptic one.

43. Landy MS, Maloney LT, Johnston EB, Young M: **Measurement and modeling of depth cue combination: in defense of weak fusion**. *Vision Res* 1995, **35**:389-412.

44. Yuille A, Grzywacz N: **A computational theory for the perception of coherent visual motion**. *Nature* 1988, **333**:71-74.

45. Landy MS, Kojima H: **Ideal cue combination for localizing texture-defined edges**. *J Opt Soc Am A* 2001, **18**:2307-2320.

46. • Saunders JA, Knill DC: **Perception of 3D surface orientation from skew symmetry**. *Vision Res* 2001, **41**:3163-3183.

The projection of a symmetric flat object has a distorted or 'skewed' symmetry that provides partial information about the object's orientation in depth. The authors show that human observers integrate symmetry information with stereo information weighted according to the reliability of the source.

47. Mamassian P, Landy MS: **Interaction of visual prior constraints**.
• *Vision Res* 2001, **41**:2653-2668.
The authors and others had previously shown that human observers interpret shape assuming prior knowledge constraints that both the viewpoint and the light source tend to be above the object. The authors manipulated the reliability of each of the two constraints by changing the contrast of different parts of the stimuli. The results suggested that prior knowledge constraints behave like depth cues in integration: that is, cues with more reliable information have higher weight attributed to their corresponding prior knowledge constraint.

48. Bülthoff HH, Mallot HA: **Integration of depth modules: stereo and shading**. *J Opt Soc Am A* 1988, **5**:1749-1758.

49. Pearl J: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* 2nd Edn. San Mateo: Morgan Kaufmann Publishers; 1988.

50. Lorenceau J, Shiffrar M: **The influence of terminators on motion integration across space**. *Vision Res* 1992, **32**:263-273.

51. McDermott J, Weiss Y, Adelson EH: **Beyond junctions: nonlocal**
• **form constraints on motion interpretation**. *Perception* 2001, **30**:905-923.
The interpretation of the translating diamond (described by [50]) shown in (Figure 4d, depends on which local motion measurements are integrated. These authors show how non-local form information related to surface occlusions determines the selection of motion elements to be integrated. The web-based interactive demonstrations are highly recommended: http://www.perceptionweb.com/perc0801/square.html

52. Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL:
•• **Shape perception reduces activity in human primary visual cortex**. *Proc Natl Acad Sci USA* 2002, **99**:15164-15169.
Earlier fMRI work by several groups (see [61]) has shown that the human lateral occipital complex (LOC) increases its activity during object perception. The authors use fMRI to show that when local visual information is perceptually organized into whole objects, activity in V1 decreases over the same period that activity in higher, LOC areas increases. The authors interpret the activity changes in terms of high-level hypotheses that compete to explain away the incoming retinal data.

53. Mamassian P, Knill DC, Kersten D: **The perception of cast shadows**. *Trends Cogn Sci* 1998, **2**:288-295.

54. Blake A, Bülthoff HH: **Does the brain know the physics of specular reflection?** *Nature* 1990, **343**:165-169.

55. Tu Z, Zhu S-C: **Image segmentation by data-driven Markov**
•• **chain**. *IEEE T Pattern Anal* 2002, **24**:657-673.
A Bayesian model for image segmentation that uses low-level visual features to rapidly search through high-level interpretations. This model yields good segmentations when tested on a database of a hundred natural images and compared with ground-truth estimates by human observers.

56. Burgi PY, Yuille AL, Grzywacz NM: **Probabilistic motion estimation based on temporal coherence**. *Neural Comput* 2000, **12**:1839-1867.

57. Lee TS, Yang CF, Romero RD, Mumford D: **Neural activity in early**
•• **visual cortex reflects behavioral experience and higher-order perceptual saliency**. *Nat Neurosci* 2002, **5**:589-597.
Electrode recording of cells in V1 and V2 of macaque monkeys respond to the apparently high-level task of detecting stimuli that pop-out from shape-from-shading. These responses changed with the animal's behavioral adaptation to contingencies, which suggests dependence on experience and utility.

58. Pouget A, Dayan P, Zemel R: **Information processing with population codes**. *Nat Rev Neurosci* 2000, **1**:125-132.

59. Mumford D: **On the computational architecture of the neocortex. II. The role of cortico-cortical loops**. *Biol Cybern* 1992, **66**:241-251.

60. Bullier J: **Integrated model of visual processing**. *Brain Res Brain*
• *Res Rev* 2001, **36**:96-107.
This paper reviews the evidence that feedback connections are used to enable high-level visual areas to influence perception by interacting with low-level areas V1 and V2. In particular, the cooling of areas such as motion area MT (middle temporal cortex) is shown to affect the response of neurons in lower-areas.

61. Grill-Spector K, Kourtzi Z, Kanwisher N: **The lateral occipital complex and its role in object recognition**. *Vision Res* 2001, **41**:1409-1422.

62. Tu Z, Zhu S-C: **Parsing images into region and curve processes**.
• In *Proceedings of the 7th European Conference on Computer Vision April 29,2002; Copenhagen, Denmark: LNCS2352*. Edited by Heyden A, Sparr G, Nielsen M, Johnasen P. Berlin, Heidelberg: Springer-Verlag; 2002:393. URL: http://link.springer.de/link/service/series/0558/tocs/t2352.htm#toc2352
This paper uses a feedforward and feedback Bayesian model to parse images into regions and curves by requiring these alternative explanations to compete to explain the data. The curves can 'explain away' image features, which might adversely affect the quality of the image segmentations by distorting the estimates of the regions.

63. Belhumeur PN, Kriegman DJ, Yuille A: **The bas-relief ambiguity**. In *IEEE Conference on Computer vision and Pattern Recognition*. Puerto Rico: IEEE Computer Society;1997: 1060-1066.

64. Sinha P, Adelson E: **Recovering reflectance and illumination in a world of painted polyhedra**. In *Proceedings of Fourth International Conference on Computer Vision*; Berlin: IEEE Computer Society Press;1993:156-163.

65. Kersten D, Schrater PR: **Pattern inference theory: a probabilistic approach to vision**. In *Perception and the Physical World*. Edited by Mausfeld R, Heyer D: Chichester: John Wiley & Sons, Ltd.; 2002: 191-228.