

# Comparison of Bayesian and empirical ranking approaches to visual perception

Catherine Q. Howe<sup>a</sup>, R. Beau Lotto<sup>b</sup>, Dale Purves<sup>a,\*</sup>

<sup>a</sup>Center for Cognitive Neuroscience and Department of Neurobiology, Duke University, Durham NC 27708, USA

<sup>b</sup>Institute of Ophthalmology, University of College London, London W13, UK

Received 22 July 2005; received in revised form 12 January 2006; accepted 18 January 2006

Available online 14 March 2006

## Abstract

Much current vision research is predicated on the idea—and a rapidly growing body of evidence—that visual percepts are generated according to the empirical significance of light stimuli rather than their physical characteristics. As a result, an increasing number of investigators have asked how visual perception can be rationalized in these terms. Here, we compare two different theoretical frameworks for predicting what observers actually see in response to visual stimuli: Bayesian decision theory and empirical ranking theory. Deciding which of these approaches has greater merit is likely to determine how the statistical operations that apparently underlie visual perception are eventually understood.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Vision; Perception; Bayesian decision theory; Empirical ranking theory

## 1. Introduction

A central problem in vision recognized for the last several centuries is that the information in retinal stimuli cannot be mapped unambiguously back onto real-world sources, a quandary referred to as the “inverse optics problem” (Berkeley, 1709/1976). The basis of this problem is straightforward. With respect to the physical characteristics of light reaching the eye from any source, illumination, reflectance and transmittance are inevitably conflated in the retinal image (Fig. 1A). Similarly, the shape and spatial arrangement of objects are uncertain because information about size, distance and orientation is also conflated in retinal stimuli (Fig. 1B). In consequence, how patterns of light on the retina are related to their generative physical sources cannot be known by direct means. Nevertheless, to be successful, visually guided behavior must deal appropriately with the physical sources of light stimuli. Thus, the uncertain relationship of retinal images

to their real-world provenance presents a profound challenge for theories of vision.

Recognition of this dilemma over the years has led to a number of speculations about how the visual system might use empirical information (i.e., past experience) to enable appropriate responses to a world that cannot be known directly. The first and most influential advocate of using past experience as a means of contending with the uncertain source of retinal stimuli was Helmholtz (1866/1924). Helmholtz summarized his conception of the empirical contribution to visual percepts by proposing that the raw “sensations” generated by the physiological infrastructure of the eye and the input stages of the visual brain could be modified by information derived from the experience of each individual. He described this process as making “unconscious inferences” about reality, thus generating perceptions more nearly aligned with stimulus sources when input-level “sensations” proved inadequate (op. cit., vol. III, p. 10 ff).

Despite Helmholtz’s stature and the prominence of ensuing arguments about these issues during the second-half of the 19th century (Turner, 1994), vision science during most of the 20th century has been understandably

\*Corresponding author. Tel.: +1 919 684 6122; fax: +1 919 684 4431.

E-mail address: [purves@neuro.duke.edu](mailto:purves@neuro.duke.edu) (D. Purves).

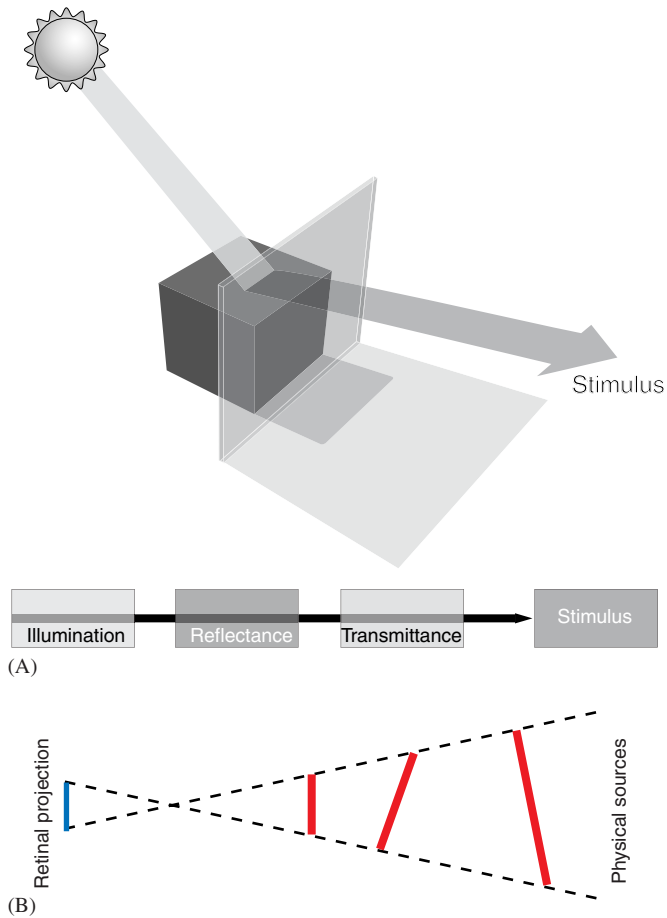


Fig. 1. A basic quandary in vision is the necessarily uncertain relationship between the information in the images that fall on the retina and their real-world sources: (A) The inherent ambiguity of the amount and quality of light falling on the retina. The fundamental factors that determine the luminance and spectral distribution of any stimulus are illumination, reflectance, and transmittance, which are always conflated in the projected image. (B) This problem is equally apparent in the spatial domain. Thus, the same retinal projection can be generated by objects of different sizes at different distances from the observer, and in different orientations.

dominated by the enormous success of modern neurophysiology and neuroanatomy. A plausible assumption in much of this research has thus been that understanding visual perception will be achieved through increasingly precise information about the receptive field properties of visual neurons and the synaptic connectivity that gives rise to these properties (see, for example, Hubel, 1988, p. 85 ff). As a result, the role of past experience in determining what observers see has, until recently, received relatively little attention.

The last few years, however, have witnessed a resurgence of interest in the role of empirical information in vision, driven primarily by vision scientists with backgrounds in psychology, computer science, statistics and mathematics (Brainard and Freeman, 1997; Freeman, 1994; Geisler et al., 2001; Heyer and Mausfeld, 2002; Kersten, 2000; Knill and Richards, 1996; Purves and Lotto, 2003; Rao et al.,

2002; Weiss et al., 2002; Yuille and Grzywacz, 1998). The uncertain provenance of retinal stimuli and the question of how the visual system contends with this quandary have thus returned to center stage.

## 2. Bayes' theorem

If the visual system uses information gleaned from past experience to overcome the inverse optical problem, then understanding vision inevitably means understanding how, in statistical terms, physical sources are related to retinal images. By far the most popular approach to sorting out this relationship has been Bayesian decision theory (e.g., Knill and Richards, 1996; Rao et al., 2002; Maloney, 2002).

Thomas Bayes was an 18th century minister and amateur mathematician whose paper in 1763 proved a theorem showing how conditional probabilities are used in making inferences (Bayes, 1763). Although Bayes' purpose in elaborating his eponymous theorem remains obscure, it has been applied in a variety of disciplines as a framework for addressing statistical problems whose solution depends on an assessment of hypotheses that are only more or less likely to be true as a result of complex circumstances. In vision research, Bayes' theorem was initially used to develop pattern recognition strategies for computer vision (e.g., Geman and Geman, 1984; Grenander, 1996). More recently, however, the framework provided by the theorem has been advocated as a means of rationalizing aspects of biological vision (Maloney, 2002; Geisler and Kersten, 2002; Kersten and Yuille, 2003; Knill et al., 1996; Mamassian et al., 2002; Knill and Pouget, 2004).

Bayes' theorem is usually written in the form

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)},$$

where  $H$  is the hypothesis,  $E$  the evidence pertinent to its validity and  $p$  the probability. The first term on the right-hand side of Bayes' equation,  $P(H)$ , is referred to as the *prior probability distribution* or simply the *prior*, and is a statistical measure of confidence in the hypothesis, absent any present evidence pertinent to its truth or falsity. With respect to vision, the prior describes the relative probabilities of different physical states of the world pertinent to retinal images, i.e., the relative frequency of occurrence of various illuminants, surface reflectance values, object sizes and so on. The second term,  $P(E|H)$ , is called the *likelihood function*. If hypothesis  $H$  were true, this term indicates the probability that the evidence  $E$  would have been available to support it. In the context of vision, given a particular state of the physical world (i.e., a particular combination of illumination, reflectance properties, object sizes, etc.), the likelihood function describes the probability that the state would generate the retinal projection in question. The product of the prior and the likelihood function, divided by a normalization constant,  $P(E)$ , gives the *posterior probability distribution*,  $P(H|E)$ . The posterior distribution defines the probability of hypothesis  $H$  being true, given

the evidence  $E$ . In vision, the posterior probability distribution thus indicates the relative probability of a given retinal image having been generated by *one or another of the different physical realities that might be the source of the image*.

To illustrate how Bayes' theorem can be used to rationalize the percepts elicited by a visual stimulus, consider the apparent brightness of a patch having a particular luminance value in the retinal image. (Luminance describes the intensity of light falling on the retina, corrected for the sensitivity of human vision). Under natural conditions, the intensity of a light stimulus is, to a first approximation, a direct product of the amount of light falling on object surfaces in the world (the illumination) and the various reflectance efficiency functions of those surfaces. To further simplify the situation, consider illumination ( $\Omega$ ) and reflectance ( $R$ ) the only parameters needed to specify the physical reality underlying the image (in reality, the amount of light reaching the retina will of course be affected by the transmittance of the intervening atmosphere (see Fig. 1A), the characteristics of the ocular media, and a host of other physical factors).

Applying Bayes' theorem, the brightness experienced by an observer in this simplistic example is predicted by the pertinent posterior probability distribution  $P(\Omega, R|L)$ , where  $L$  stands for the luminance of the image. In making this prediction, the first step is to determine the prior distribution  $P(\Omega, R)$ , i.e., the probability distribution of different conditions of illumination occurring in conjunction with different surface reflectance values in the physical world (Fig. 2A). In principle, this distribution could be generated by sampling illumination and surface reflectance at a large number of points in typical physical environments (a difficult but not impossible task; in practice, illumination and reflectance values are usually assumed). The next step is to derive the likelihood function  $P(L|\Omega, R)$ , which describes, for each possible combination of illumination and reflectance, the probability that the combination would generate the luminance value of the image patch in question. If this process were free of noise, then particular combinations of illumination and reflectance would always generate a specific value of luminance. With respect to the luminance of any given image, some of these combinations would inevitably produce that particular luminance and

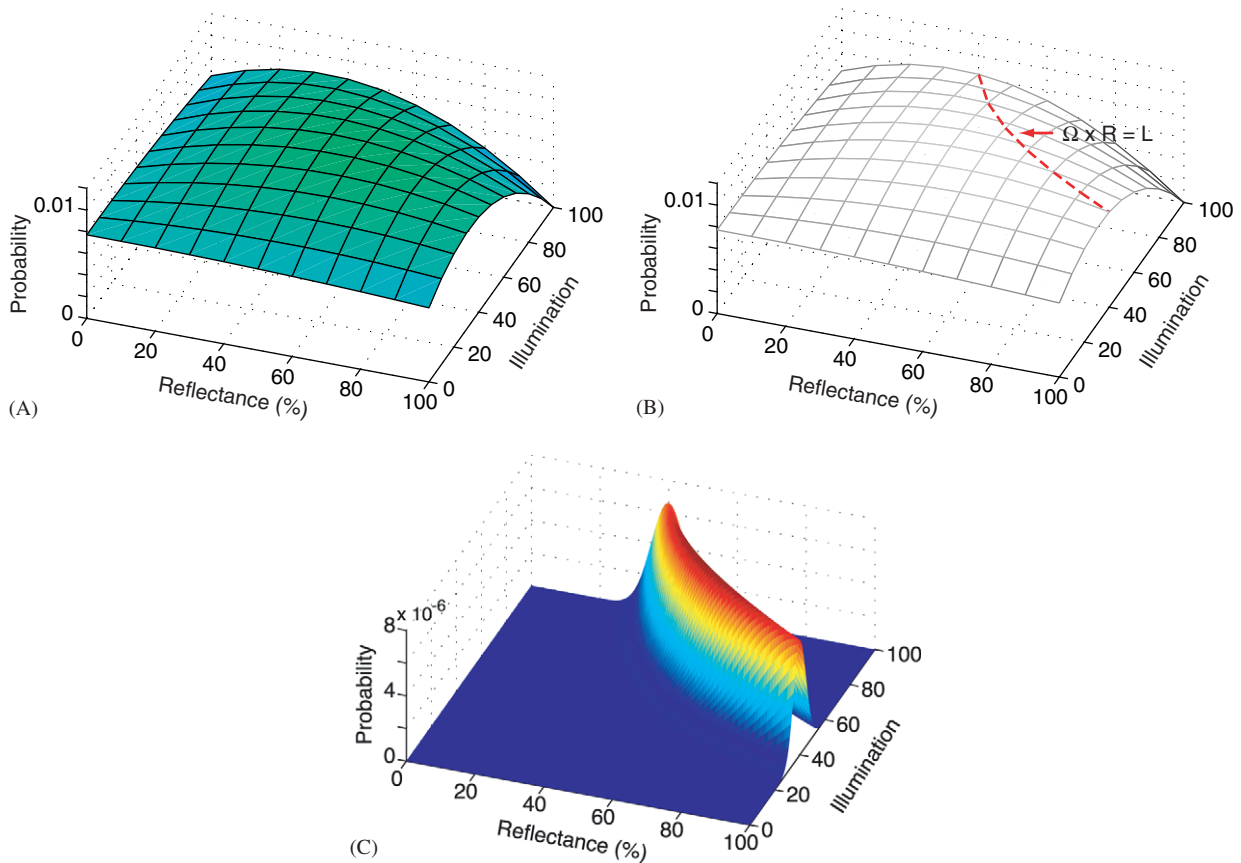


Fig. 2. Bayesian approach to characterizing the relationship between the luminance of a visual target and its possible physical sources: (A) A prior distribution of illuminant ( $\Omega$ ) and reflectance values ( $R$ ) in the physical world. The distribution, which is didactic only, shows illumination on an arbitrary scale from 0 to 100, and reflectance on a scale from 0% to 100%. (B) The dashed red line on the prior distribution indicates the set of points at which the product of illumination and reflectance equals the luminance of the target ( $L$ ). If the image formation process is assumed to be free of noise, the posterior distribution,  $P(\Omega, R|L)$ , obtained by multiplying the prior by a likelihood function will be a section of the prior indicated by dashed line. (C) The addition of Gaussian noise to the image formation process makes the posterior distribution “thicker”, but does not alter the fact that the posterior is effectively a section of the prior (color indicates the relative probability values).

therefore have a probability of 1 in the likelihood function, whereas all others would be incapable of producing that luminance, and thus have a probability of 0. Because biological image formation is inevitably noisy, some Gaussian noise is typically added in Bayesian models, making the values of 1 in the likelihood function somewhat less than 1, and values of 0 somewhat larger than 0.

Finally, the posterior distribution,  $P(\Omega, R|L)$ , is obtained by multiplying the prior distribution in Fig. 2A by the likelihood function (which comprises zeros and ones, or, with added noise, approximations thereof), and dividing by the normalization factor in Bayes' theorem. The posterior distribution therefore describes the relative probabilities of all the possible physical sources that could have generated the specific image under consideration, and is effectively a section of the prior distribution. In the example here the section is the set of all the specific combinations of illumination and reflectance values whose product is  $L$  (Fig. 2B). The addition of Gaussian or other noise to the likelihood function only increases the "thickness" of the section (Fig. 2C).

### 3. Bayesian decision theory

Because the posterior distribution indicates only the relative probabilities of a set of possible image sources, a particular source (i.e., a particular combination of illumination and reflectance in the example above) must be selected from this set if the aim is to predict what brightness an observer will actually see. The usual way of addressing this further requirement is to assume that the visual system makes this "choice" according to the behavioral consequences associated with each perceptual "decision". The influence of behavioral consequences is typically expressed in terms of the discrepancy between the decision made and the actual state of the world, which over the full range of the possible choices defines a *gain-loss* function. Since there is no a priori way to model this function (indeed, given the enormous number of variables involved, a realistic gain-loss function for some aspect of vision would be extraordinarily difficult to determine), the relative cost of different behavioral responses is also assumed. For example, a common assumption is that observers will "choose" the percept that corresponds to the maximum value in posterior probability distribution, since this choice would generally minimize the discrepancy between the percept and the actual state of the world.

The requirement of a gain-loss function in using Bayes' theorem to predict visual percepts explains why this general approach is referred to as *Bayesian decision theory*. In this framework, the *specific physical condition* thus determined corresponds to what an observer should see in response to the stimulus in question. Thus the brightness seen in response to the luminance of the stimulus in the example here corresponds to the most probable physical reflectance and illumination values underlying the stimulus, much as Helmholtz initially suggested. (Helmholtz and others

concluded that observers do not perceive illumination as such, and thus interpret the resulting percept as seeing the most likely underlying reflectance.)

In sum, Bayesian decision theory determines the physical source(s) capable of generating a given retinal image and the relative probabilities of their actually having done so according to the observer's experience; the percepts predicted are therefore "explicit models of world structure" (Knill et al., 1996, p. 7). To date, this approach has been used to rationalize several aspects of visual perception, including the perception of surface orientation and shape based on texture, shading and motion (Blake et al., 1993; Jacobs, 1999; Knill, 1998a b), the perceptual organization of contours (Geisler et al., 2001; Feldman, 2001) and a variety of motion illusions (Weiss et al., 2002).

### 4. Empirical ranking theory

The application of Bayesian decision theory to vision is an important advance in that it formalizes Helmholtz's qualitative proposal about "visual inferences" as a means of contending with the inevitably uncertain sources of visual stimuli. Its implementation, however, presents both conceptual and practical difficulties. With respect to conceptual limitations, the intuitively appealing idea that percepts correspond to physical characteristics, such as surface reflectance, is problematic and sometimes demonstrably false (as explained below). Practical obstacles, as noted, are the difficulty of measuring the physical parameters relevant to any specific prior, the problem of determining likelihood functions in natural settings, and the need for a decision rule based on an assumed gain-loss function. Is there, then, another way of conceptualizing how vision uses empirical information to deal with the inverse optics problem, and of predicting what people should see in response to various retinal stimuli?

The alternative is to simply abandon the long-held assumption that vision entails inferences about the properties of the physical world, and that what we see corresponds to these properties. In this approach, which we have called *empirical ranking theory*, the perceptual quality elicited by any particular aspect of a retinal stimulus (e.g., the brightness elicited by the luminance of a given retinal image) is predicted by the relative frequency of occurrence of that parameter—the luminance of the patch in this example—in relation to all the other instances of luminance that have been experienced in the same context (Yang and Purves, 2004).

Empirical ranking thus differs fundamentally from the Bayesian idea that percepts correspond to the *most likely physical source of a stimulus*. In the example in Fig. 2, for instance, the brightness elicited by the luminance of a given image patch is predicted by the empirical rank of the relevant luminance value within the full range of experience with the luminance of such patches in similar scenes. This way of generating subjective experience allows the full range of percepts pertinent to a given visual quality

(e.g., percepts that range from the brightest to the dimmest) to be aligned with the full range of the relevant stimulus parameters generated by the physical world (from the most intense luminance to the least intense).

To illustrate the approach in empirical ranking more specifically, consider the apparent brightness of a patch in the retinal image. In this framework, the relative frequency of different states of the physical world, represented by the prior distribution in Fig. 2A, is integrated according to the luminance values produced by every combination of illumination and reflectance in the distribution (Fig. 3A), in this way generating a marginal (i.e., integrated) form of the prior (Fig. 3B). The marginal distribution in Fig. 3B thus describes the probabilities of occurrence of the full

range of possible luminance values, each point in the distribution being the summed probability of occurrence of all the physical conditions that could have generated a specific luminance. In practice, this information can be determined simply by measuring the relative frequency of luminance values in natural images (Yang and Purves, 2004), a sufficient number of which serves as a proxy for human visual experience over both evolutionary time and the lifespan of an individual.

A cumulative probability distribution can then be derived from the marginal distribution that is effectively an empirical scale that orders the full range of luminance values in terms of past experience (Fig. 3C). The rank of any given luminance on the scale is determined by the percentage of all the physical sources that generated luminance values less than the value at issue, and the percentage that generated greater luminance values. The higher the percentage of physical sources that in past experience generated luminance values less than the luminance in question, the higher that luminance ranks on the empirical scale defined by the cumulative distribution, and thus the brighter the percept elicited, as illustrated in Fig. 3C. In short, such ranking maps the full range of a stimulus feature (luminance in this case) to the full range of the corresponding perceptual space (brightness) according to past experience.

Importantly, the rank of any luminance—and thus the sensation of brightness that is ultimately perceived—bears no direct relationship to the possible underlying values of reflectance and illumination in the physical world, taking on meaning only in comparison to the rank of other luminance values on the same empirical scale. As a result, the perceptual prediction generated by empirical ranking always entails an assessment of two or more stimuli, or two or more regions within a stimulus. The same general point applies to any perceptual quality, consistent with the fact that visual perception fails in the presence of a truly uniform visual field (a so-called “ganzfeld”). Notice further

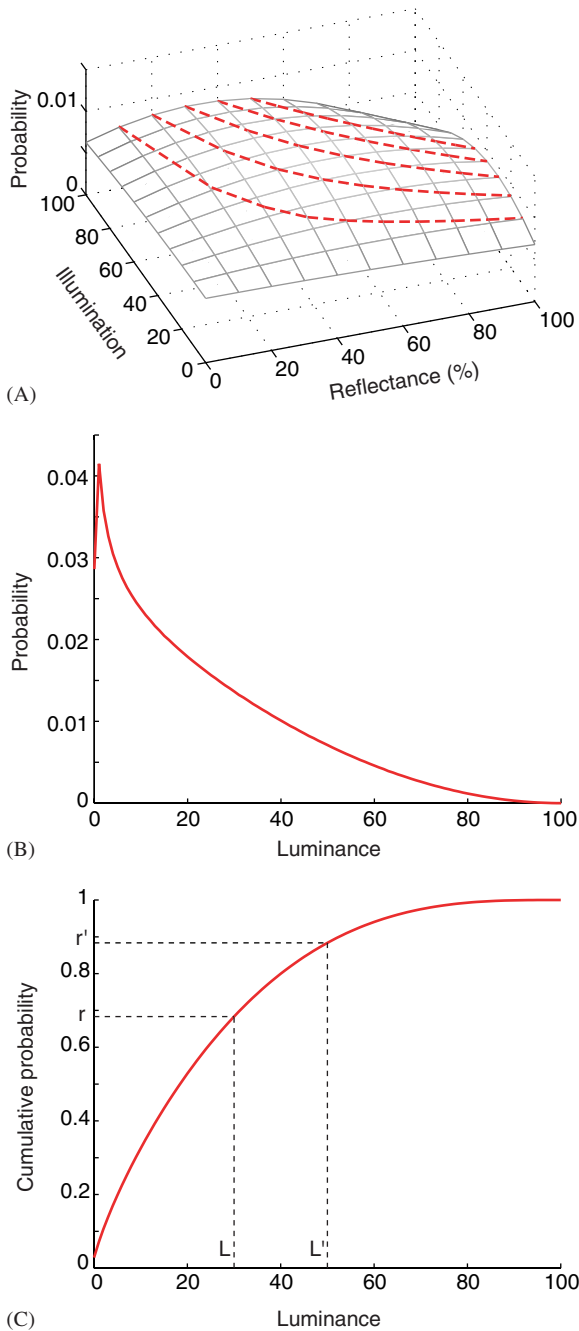


Fig. 3. Empirical ranking approach to characterizing the relationship between the luminance of a target and its possible physical sources: (A) The prior distribution in Fig. 2A is integrated along a series of curves (dashed lines) to produce the marginal distribution shown in (B). Each of these curves is an iso-luminant line along which the product of illumination and reflectance has a specific luminance value. (B) The marginal distribution derived by integrating the distribution in (A) describes the relative probability of occurrence of the physical sources of different luminance intensities in human experience. (C) The cumulative probability distribution derived from (B). The cumulative probability for any specific luminance value,  $l$ , is the summed probability of occurrence of the physical sources that have generated luminance values less than or equal to that luminance, and is derived by calculating the area underneath the function in (B) and to the left of the position where  $x = l$ . The  $y$ -value of each point on the cumulative distribution thus indicates the percentage of physical sources that generate luminance values less than or equal to a specific luminance, providing a basis for ranking that luminance among other luminance values experienced in the past. In the example shown, luminance  $L'$  holds a higher rank ( $r'$ ) than luminance  $L$  (which holds rank  $r$ ), and should thus be seen as brighter than luminance  $L$ .

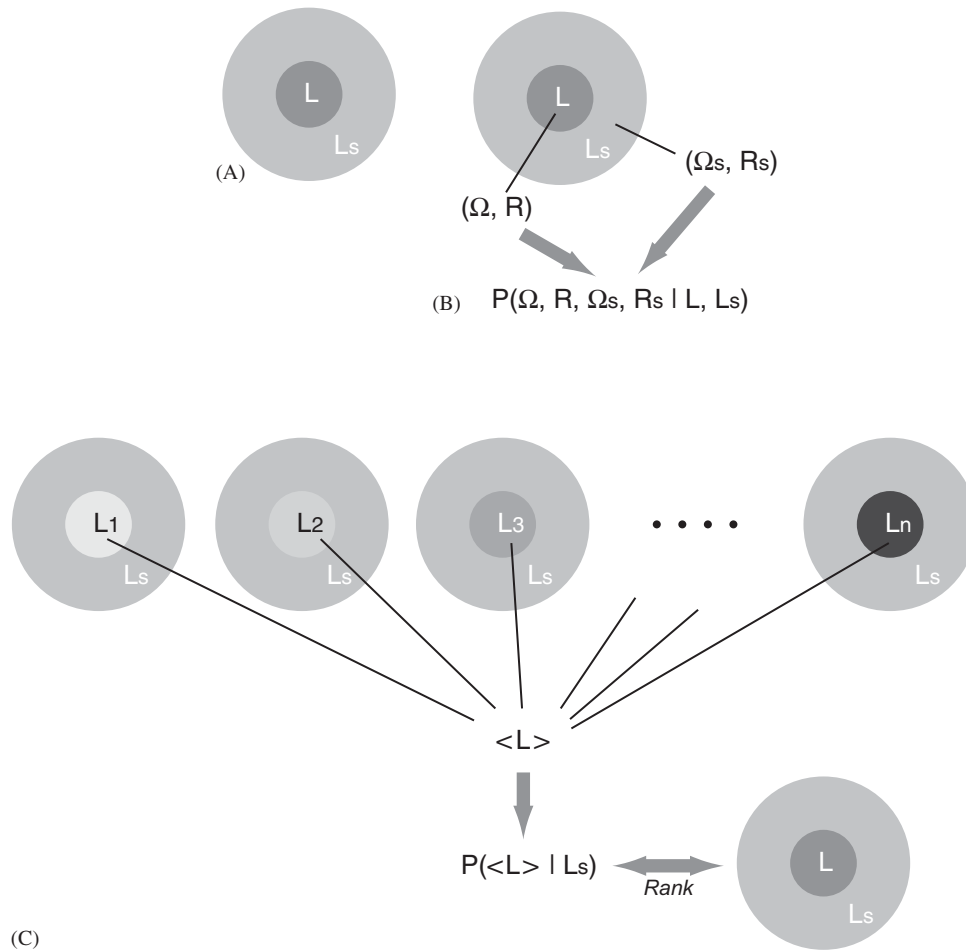


Fig. 4. Comparison of Bayesian and empirical ranking approaches to understanding contextual effects: (A) An image patch of a given luminance ( $L$ ) embedded in the surrounding area that has a different luminance value ( $L_s$ ). (B) A Bayesian approach entails the determination of the posterior distribution, which as indicated describes the relative probabilities of the sets of illumination and reflectance values ( $\Omega$ ,  $R$ ;  $\Omega_s$ ,  $R_s$ ) that could have generated the luminance value of the target and the surround, respectively. (C) The empirical ranking approach entails the determination of the probability distribution  $P(\langle L \rangle | L_s)$ , which indicates the relative probabilities of the possible luminance values of a central patch co-occurring with the specific luminance value ( $L_s$ ) of the surround. This distribution provides an empirical scale that ranks any specific target luminance (such as the luminance value  $L$ ).

that because there is no equivalent of a likelihood function or posterior distribution in this formulation, empirical ranking is conceptually and computationally distinct from Bayesian decision theory. Like Bayesian analysis, however, empirical ranking can incorporate any number of co-occurring parameters, making it useful for multi-dimensional analysis as well investigation of a single parameter (e.g., Long et al., 2006).

## 5. The importance of context

In nearly all natural visual scenes any given image patch is embedded in a surrounding context; as a result, the brightness of a patch, or any perceived target quality, will always be influenced by the surround of the target. How, then, do these two different approaches—the Bayesian decision theory and the empirical ranking theory—explain contextual influences, and how successful are they in meeting this challenge?

To answer this question, consider the patch of luminance  $L$  already described, but now surrounded by an area that has another luminance value  $L_s$  (Fig. 4A). In rationalizing the influence of the surround, the Bayesian approach would entail obtaining the relevant posterior distribution  $P(\Omega, R, \Omega_s, R_s | L, L_s)$ , where  $\Omega_s$  and  $R_s$  are the illumination and reflectance values of the possible physical conditions giving rise to the area surrounding the target patch (Fig. 4B). The posterior thus describes the relative probabilities of the various sets of real-world illumination and reflectance values that could generate, respectively, the luminance values of the target *and* the surround. As before, applying a decision rule to the posterior based on an assumed gain–loss function would then generate the specific illumination and reflectance values taken to predict the perception of brightness elicited by the target in conjunction with its surround.

In contrast, empirical ranking theory computes the probability distribution  $P(\langle L \rangle | L_s)$ , where  $\langle L \rangle$  represents

the full range of the possible luminance values of a target patch that have occurred in past experience in conjunction with the specific luminance value ( $L_S$ ) of the surround (Fig. 4C). Like the marginal probability distribution in Fig. 3B, the distribution  $P(\langle L \rangle | L_S)$  provides an empirical scale that ranks *all* the target luminance values that have co-occurred with the surround in question. Thus, for any particular target luminance ( $L$ ), this distribution indicates, among all the physical sources of a target patch that have co-occurred with the surround, the percentage that generated target patches having luminance values less than  $L$ , and the percentage that generated target patches with luminance values greater than  $L$ . As before, the higher  $L$  ranks on this empirical scale, the brighter the perception of the target patch.

The example of the luminance patch as used in the preceding section in which the surround was not considered might have made it seem that predicting the brightness elicited by different luminance values according to their empirical rank is trivial in that a higher luminance value will always have a higher rank than a lower value. However, once context is introduced, the relationship between a luminance value and its empirical rank is no longer so simple. Because the empirical scales associated with various surrounds can be quite different, the same target luminance value can have very different ranks when the targets are embedded in different contexts, and can thus generate quite different perceptions of brightness.

## 6. Application of the two theories to explaining a simple contextual effect

The simplest example of the effect of context on the brightness of a target is a standard simultaneous brightness contrast stimulus. As illustrated in Fig. 5A, when two identical target patches are embedded in surrounds with different luminance values, the target in the lighter surround appears darker than the one in the darker surround.

A Bayesian approach to understanding this contextual effect would again begin by determining the posterior distribution  $P(\Omega, R, \Omega_S, R_S | L, L_S)$  for the two identical targets and the different surrounds. Although to our knowledge the issue has never been examined, it seems safe to assume that an analysis of the illumination and reflectance values underlying such an image pattern in natural settings would show that the average illumination of the target patch in the lighter surround is higher than the average illumination of the target in the darker surround, and that the average surface reflectance of the target in the lighter surround is lower than that of the target in the darker surround. Accordingly, the target patch embedded in the higher luminance surround would be more likely to have been generated by a relatively high level of illumination in conjunction with a relatively low reflectance surface patch, whereas the target embedded in a low luminance surround would be more likely to have been generated by a

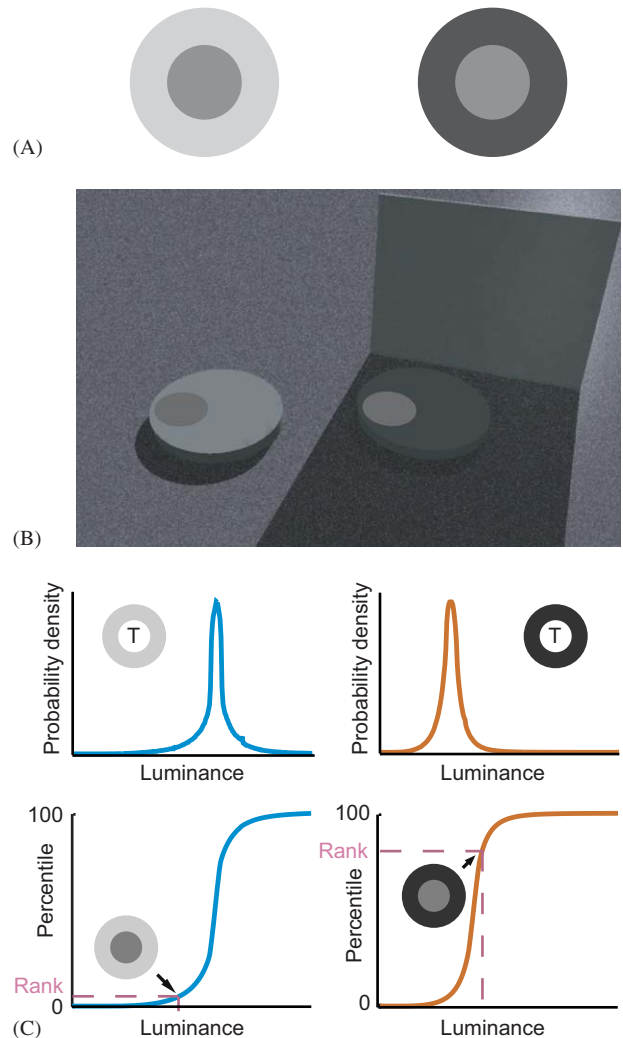


Fig. 5. Bayesian and empirical ranking explanations of a standard simultaneous brightness contrast effect: (A) Although the left and right central targets in this illustration have the same luminance, the target embedded in the surround having a relatively high luminance (left) appears darker than the target embedded in the lower luminance surround (right). (B) In Bayesian terms, the most likely physical source of the target in the higher luminance surround is a relatively low reflectance surface in relatively intense illumination, and conversely for the target on the right. (C) In empirical ranking theory, the higher luminance surround co-occurs more frequently with relatively high luminance targets, and conversely for the darker surround (upper panels;  $T$ , target). In these terms, the perceptual effect in (A) arises because a given target luminance ranks lower on the empirical scale associated with the lighter surround than on the scale associated with the darker surround, as indicated by the dashed lines in the cumulative distributions in the lower panels. (C is after Yang and Purves, 2004).

relatively low level of illumination and a relatively high surface reflectance (Fig. 5B). As a result, the most likely surface reflectance underlying the target embedded in the relatively high luminance surround would be less than the most probable surface reflectance underlying the same target in the relatively low luminance surround. If one accepts Helmholtz's speculation that illumination is discounted (see earlier), and that the brightness we see

therefore corresponds to surface reflectance, then a Bayesian approach correctly predicts that the target in the lighter surround in Fig. 5A appears darker than the target in the darker surround.

Empirical ranking theory explains such simultaneous brightness contrast effects on a different basis. When the marginal distribution  $P(\langle L \rangle | L_S)$  is derived from a database of natural images for each of the target-surround combinations in Fig. 5A, it is apparent that the relatively high luminance surround co-occurs more frequently with high-luminance targets than with low-luminance targets, and that the relatively low luminance surround co-occurs more frequently with low-luminance targets (Yang and Purves, 2004). Thus, a given target luminance associated

with a high-luminance surround ranks relatively low on the empirical scale of target luminance values, whereas the same target luminance ranks relatively high on the empirical scale associated with a low-luminance surround (Fig. 5C). Accordingly, the target embedded in the relatively high luminance surround in Fig. 5 should appear darker than the same target in the low luminance surround, as it does. Thus, for a relatively simple contextual effect both approaches predict the perceptual effect.

### 7. Explaining the effects of more complex contexts

Such simple stimuli, however, do not exemplify the far more complicated contrast patterns found in natural

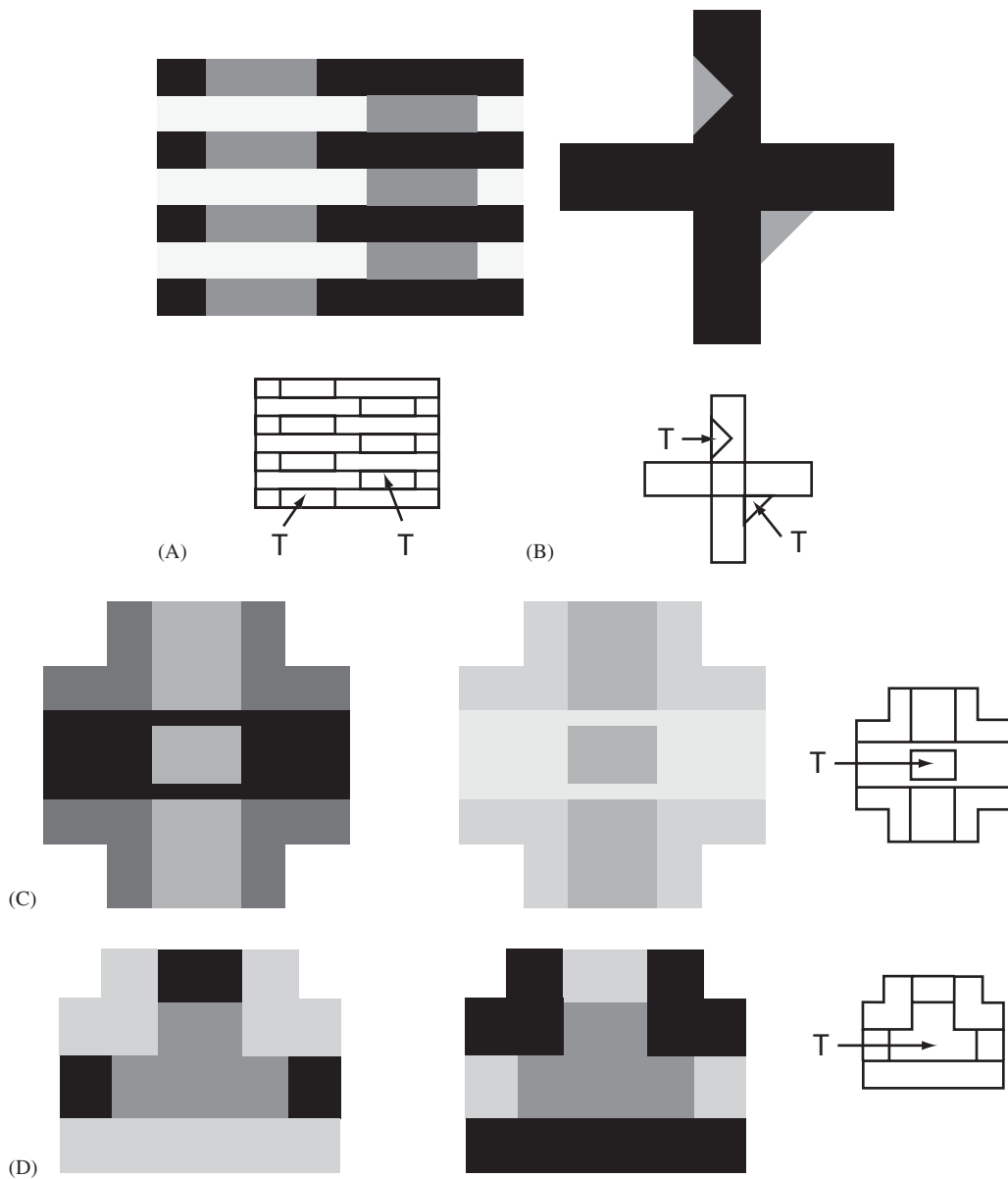


Fig. 6. The perceptual effects elicited by more complex luminance patterns. The targets (*T*) in each stimulus are equiluminant, but appear different in brightness depending on the type of luminance pattern in which they appear: (A) White's illusion. (B) Wertheimer–Benary illusion. (C) The intertwined cross illusion. (D) The inverted T-illusion. (After Yang and Purves, 2004.)

scenes. By the same token, they are not particular demanding tests of either theoretical approach. Examples of more complex stimuli in the literature whose perceptual consequences provide a much greater predictive challenge include White's illusion, the Wertheimer–Benary effect, the intertwined cross illusion and the inverted-T illusion (Fig. 6).

In contrast to a simple stimulus such as that in Fig. 5A, explaining these more complex effects in terms of Bayesian decision theory is problematic. The reason a Bayesian framework correctly predicts the simultaneous brightness contrast effect elicited by a simple stimulus is that a surface patch of a given luminance in a high-luminance context will, on average, have a lower reflectance than a patch of the same luminance surrounded by a low-luminance context. For a pattern of luminance values like White's stimulus, however, the most probable reflectance values of the target patches, given their respective contexts, predict brightness percepts that are actually the opposite of those elicited in an observer. The reason is that in this case, the targets on the left in White's stimulus are embedded in a context with a higher average luminance than the context of the targets on the right. Thus, based on the argument laid out in the previous section, the physical source of the target on the left should have a lower reflectance, on average, than the source of the identical target on the right; nonetheless, the target on the left now looks brighter/lighter than the one on the right (see Fig. 6A). The same sort of difficulty arises for each of the other stimuli illustrated in Fig. 6.

Because there are so many variations of Bayesian analysis in the modern vision literature, a natural question is whether some adjustment of this theoretical framework could ultimately predict the perceptual effects elicited by these more complex stimuli. There are two basic ways a Bayesian model can be adjusted: (1) increase or decrease the number of dimensions in the prior distribution; and (2) change the way perceptual estimates are derived from the posterior distribution (i.e., adjust the gain–loss function). Adjusting these parameters, however, would not change the fact that estimates of the surface reflectances underlying the targets in Fig. 6 contradict the perceptions of brightness/lightness that they elicit.

In contrast to the problems that arise using a Bayesian framework, the percepts elicited by each of the configurations in the Fig. 6 are correctly predicted by the empirical rank of the target luminance in these patterns, determined by tallying up the relative frequency of co-occurring luminance relationships in natural scenes (Yang and Purves, 2004). Empirical ranking theory has also been able to explain equally challenging phenomena in the perception of color (Long et al., 2006) and form (Howe and Purves, 2004, 2005a–c). The greater success of this approach implies that an orderly mapping of perceptual space to physical space is the key to understanding visual percepts and the neural processing that gives rise to them; conversely, making inferences about the physical properties

of the objects giving rise to a visual stimulus, however intuitively appealing, appears to be a misleading way of thinking about how vision operates.

## 8. Conclusions

Recent research has implied that as a means of dealing with the inverse optics problem, visual percepts are generated according to the statistical prevalence of the real-world sources of light stimuli rather than the characteristics of the stimuli as such. At present, two different theoretical approaches have been used to rationalize visual percepts in a manner that takes account of the empirical nature of vision: Bayesian decision theory, which is widely used, and empirical ranking theory, which has been proposed more recently. The fundamental difference between these general frameworks is their different conception of visual perception. Bayesian decision theory, as it has been applied to vision, supposes that perceptions are effectively inferences about the physical properties of the objects or conditions underlying visual stimuli. In contrast, empirical ranking theory supposes that visual percepts simply represent an ordering of visual stimuli (brighter vs. dimmer, larger vs. smaller) according to the accumulated past experience with all such stimuli. Deciding which of these theoretical frameworks is a better guide for ultimately understanding the neural basis of what we see will ultimately depend on the relative ability of two approaches to explain the full range of the many unresolved perceptual puzzles that vision presents. For the reasons stated, empirical ranking has both conceptual and practical advantages over Bayesian decision theory as a means of understanding vision.

## Acknowledgments

We are grateful to Roland Baddeley, Jean-Marc Fellous, Bill Geisler, Fuhui Long, Larry Maloney, David Schwartz, Jim Voyvodic and Zhiyong Yang for helpful criticism, which was often spirited and not necessarily in agreement with the arguments presented here.

## References

- Bayes, T., 1763. An essay toward solving a problem in the doctrine of chances. *Philos. Trans. R. Soc.* 53, 370–418.
- Berkeley, G., 1709/1976. *A New Theory of Vision*, Everyman's Library
- Blake, A., et al., 1993. Shape from texture: ideal observers and human psychophysics. *Vision Res.* 33 (12), 1723–1737.
- Brainard, D.H., Freeman, W.T., 1997. Bayesian color constancy. *J. Opt. Soc. Am. A Optics, Image Sci. Vision* 14 (7), 1393–1411.
- Feldman, J., 2001. Bayesian contour integration. *Perception Psychophys.* 63 (7), 1171–1182.
- Freeman, W.T., 1994. The generic viewpoint assumption in a framework for visual perception. *Nature* 368 (6471), 542–545.
- Geisler, W.S., Kersten, D., 2002. Illusions, perception and Bayes. *Nat. Neurosci.* 5 (6), 508–510.
- Geisler, W.S., et al., 2001. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41 (6), 711–724.

- Geman, S., Geman, D., 1984. Stochastic relaxation Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721–741.
- Grenander, U., 1996. *Elements of Pattern Theory*. Johns Hopkins University Press, Baltimore, MD.
- Helmholtz, H.L.F.v., 1866/1924. *Helmholtz's Treatise on Physiological Optics*. The Optical Society of America.
- Heyer, D., Mausfeld, R. (Eds.), 2002. *Perception and the Physical World: Psychological and Philosophical Issues in Perception*. Wiley, New York.
- Howe, C.Q., Purves, D., 2004. Size contrast and assimilation explained by the statistics of natural scene geometry. *J. Cogn. Neurosci.* 16 (1), 90–102.
- Howe, C.Q., Purves, D., 2005a. *Perceiving Geometry: Geometrical Illusions Explained by Natural Scene Statistics*. Springer, Berlin.
- Howe, C.Q., Purves, D., 2005b. Natural scene geometry predicts the perception of angles and line orientation. *Proc. Natl Acad. Sci. USA* 102 (4), 1228–1233.
- Howe, C.Q., Purves, D., 2005c. The Müller–Lyer illusion explained by the statistics of image–source relationships. *Proc. Natl Acad. Sci. USA* 102 (4), 1234–1239.
- Hubel, D.H. (1988). *Eye, Brain and Vision*. WH Freeman, New York.
- Jacobs, R.A., 1999. Optimal integration of texture and motion cues to depth. *Vision Res.* 39 (21), 3621–3629.
- Kersten, D., 2000. High-level vision as statistical inference. In: Gazzaniga, M.S. (Ed.), *The New Cognitive Neurosciences*. The MIT Press, Cambridge, MA, pp. 353–363.
- Kersten, D., Yuille, A., 2003. Bayesian models of object perception. *Curr. Opin. Neurobiol.* 13 (2), 150–158.
- Knill, D.C., 1998a. Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Res.* 38 (11), 1683–1711.
- Knill, D.C., 1998b. Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Res.* 38 (11), 1655–1682.
- Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719.
- Knill, D.C., Richards, W. (Eds.), 1996. *Perception as Bayesian Inference*. Cambridge University Press, Cambridge.
- Knill, D.C., et al., 1996. Introduction: a Bayesian formulation of visual perception. In: Knill, D.C., Richards, W. (Eds.), *Perception as Bayesian Inference*. Cambridge University Press, Cambridge, pp. 1–21.
- Long, F., Yang, Z., Purves, D., 2006. Spectral statistics in natural scenes predict hue, saturation and brightness. *Proceedings of the National Academy of Sciences of the United States of America* (in press).
- Maloney, L.T., 2002. Statistical decision theory and biological vision. In: Heyer, D., Mausfeld, R. (Eds.), *Perception and the Physical World: Psychological and Philosophical Issues in Perception*. Wiley, New York, pp. 145–189.
- Mamassian, P., et al., 2002. Bayesian modelling of visual perception. In: Rao, R.P.N., et al. (Eds.), *Probabilistic Models of the Brain: Perception and Neural Function*. The MIT Press, Cambridge, MA, pp. 13–36.
- Purves, D., Lotto, B., 2003. *Why We See What We Do: An Empirical Theory of Vision*. Sinauer.
- Rao, R.P.N., et al. (Eds.), 2002. *Probabilistic Models of the Brain: Perception and Neural Function*. The MIT Press, Cambridge, MA.
- Turner, R.S., 1994. In *the Eye's Mind: Vision and the Helmholtz-Hering Controversy*. Princeton University Press, Princeton, NJ.
- Weiss, Y., et al., 2002. Motion illusions as optimal percepts. *Nat. Neurosci.* 5 (6), 598–604.
- Yang, Z., Purves, D., 2004. The statistical structure of natural light patterns determines perceived light intensity. *Proc. Natl Acad. Sci. USA* 101 (23), 8745–8750.
- Yuille, A.L., Grzywacz, N.M., 1998. A theoretical framework for visual motion. In: Watanabe, T. (Ed.), *High-level Motion Processing: Computational, Neurobiological, and Psychophysical Perspectives*. The MIT Press, Cambridge, MA, pp. 187–211.