In R. P. N. Rao, B. A. Olshausen & M. S. Lewicki (Eds.) (2002). Probabilistic Models of the Brain: Perception and Neural Function (pp. 13-36). Cambridge, MA: MIT Press.

1 Bayesian Modelling of Visual Perception

Pascal Mamassian, Michael Landy and Laurence T. Maloney

Introduction

Motivation

Through perception, an organism arrives at decisions about the external world, decisions based on both current sensory information and prior knowledge concerning the environment. Unfortunately, the study of perceptual decision-making is distributed across sub-disciplines within psychology and neuroscience. Perceptual psychologists and neuroscientists focus on the information available in stimuli, developmental researchers focus on how knowledge about the environment is acquired, researchers interested in memory study how this knowledge is encoded, and other cognitive psychologists might be primarily interested in the decision mechanisms themselves. Any perceptual task evidently involves all of these components and, at first glance, it would seem that any theory of perceptual decision making must draw on heterogeneous models and results from all of these areas of psychology.

Many researchers have recently proposed an alternative (see chapters in [8]). They suggested that Bayesian Decision Theory (BDT) is a convenient and natural framework that allows researchers to study all aspects of a perceptual decision in a unified manner. This framework involves three basic components: the task of the organism, prior knowledge about the environment, and knowledge of the way the environment is sensed by the organism [6]. In this chapter, we summarize the key points that make the Bayesian framework attractive as a framework for the study of perception and we illustrate how to develop models of visual function based on BDT.

We emphasize the role played by prior knowledge about the environment in the interpretation of images, and describe how this prior knowledge is represented as prior distributions in BDT. For the sake of terminological variety, we will occasionally refer to prior knowledge as "prior beliefs" or "prior constraints", but it is important to note that this prior knowledge is not something the observer need be aware of. Yet, as we shall see, these implicit assumptions can be revealed through psychophysical experimentation.

To introduce the Bayesian approach, we illustrate how to model a simplified problem of three-dimensional perception. The problem is not realistic but our intent in presenting it is to introduce the terminology, concepts, and methods of Bayesian modeling. Following the example, we illustrate how the framework can be used to model slightly more realistic problems concerning the perception of shape from shading and from contours. We conclude with a general discussion of the main issues of the Bayesian approach. Other tutorials on Bayesian modelling of visual perception with more technical details include Knill, Kersten & Yuille [7], Yuille & Bülthoff [18], and Maloney [9].

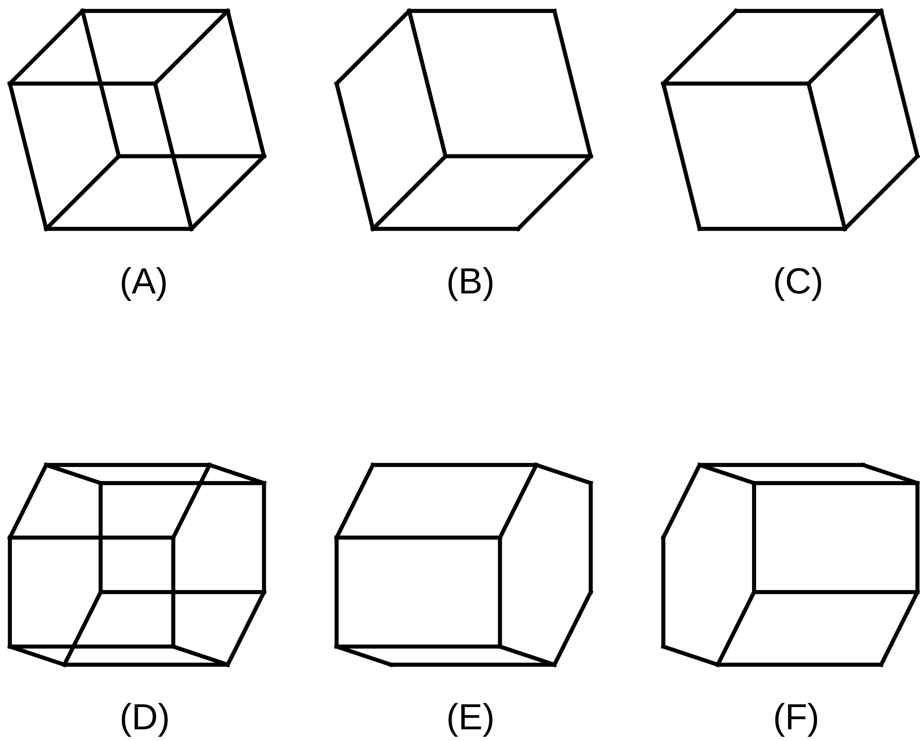

Figure 1.1: Multistable figures. Figures A and D typically engender one of two interpretations (B or C for A, and E or F for D). When viewed long enough, the interpretation will alternate between B and C and between E and F. The interpretation might vary when the figure is repeatedly presented or when the observer blinks or changes fixation.

Before introducing Bayesian modelling, we first remind the reader what makes three-dimensional perception difficult.

Visual perception as an ill-posed problem

Consider the line drawing in Fig. 1.1A. Even though this line drawing is consistent with an infinite number of polyhedra, human observers usually report only two distinct interpretations, illustrated in Figs. 1.1B and C. Similarly, Fig. 1.1D is usually interpreted as shown in Figs. 1.1E or F. If Fig. 1.1A is viewed long enough, the observer will usually experience alternation between the interpretations suggested in Figs. 1.1B and C, but only one interpretation is seen at any point in time.

Two questions arise: (1) Why are only two of an infinite number of possible interpretations seen? (2) And why do we see only one interpretation at a time rather than, say, a superposition of all of the possible interpretations?

A proper answer to the first question requires that we explain why the visual system favors certain interpretations over others *a priori*. Within the Bayesian framework, this mechanism will prove to be the prior distribution. An answer to the second question will lead us to consider how perceptual decisions are made within the Bayesian framework.

Because space perception is ill-posed with many possible scenes consistent with the available sensory information, Bayesian modelling will prove to be particularly appropriate (for a recent review of visual space perception, see [5]).

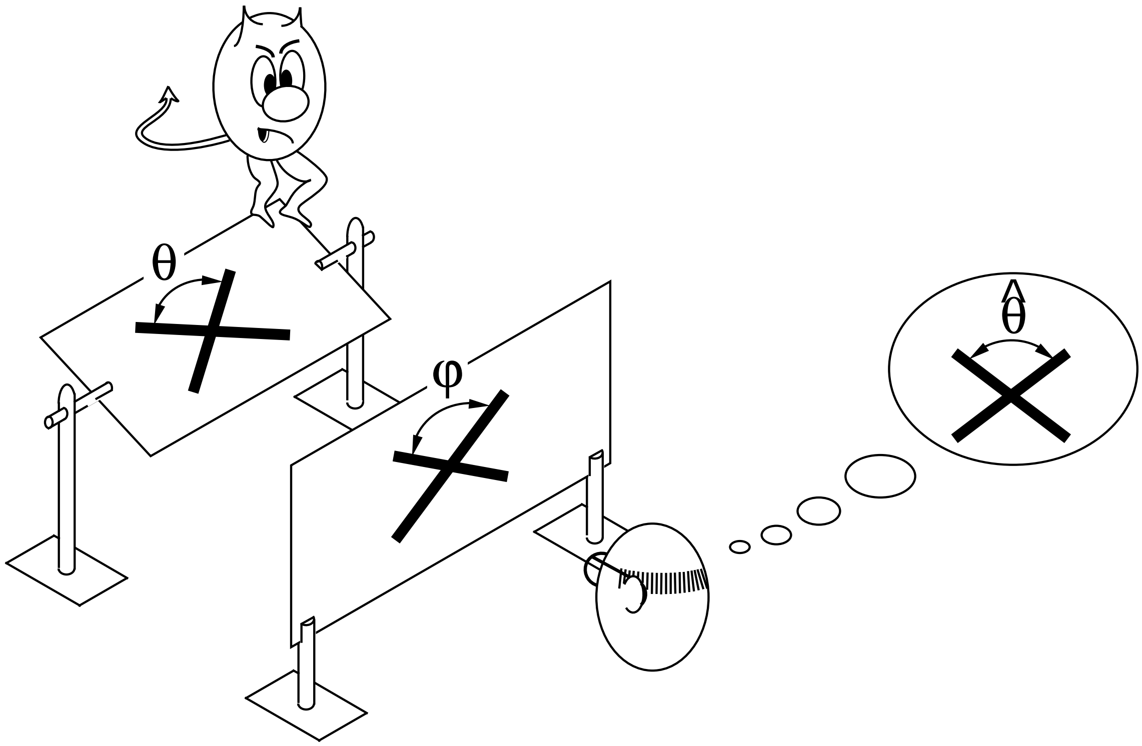

Figure 1.2: Perception of 3D angles. The angle θ between two line segments in 3D space projects as an angle φ in the image. A demon spins the physical stimulus about a horizontal axis, resulting in a random value of slant. The task of the observer is to infer θ given φ.

A Simple Example: 3D Angle Perception

In this section we introduce the key concepts of Bayesian modelling by means of a simple example. We will not attempt here to model the interpretation of the line drawings in Fig. 1.1. That would require too much investment in mathematical notation (to represent figures made out of lines in space) and would serve to obscure rather than display the elements of Bayesian modelling. Instead, we consider a perceptual task that is extremely simple, rather unrealistic, but plausibly related to perceptual interpretation of the drawings in Fig. 1.1: estimation of the angle formed by two intersecting lines in space given only the projection of the lines onto a single retina.

Task

The problem is illustrated in Fig. 1.2: Given two intersecting lines in an image forming angle φ, what is the angle θ between the two lines in three-dimensional space? Just as for Fig. 1.1, the observer looks at a two-dimensional image and has to make an inference about some three-dimensional property of the scene.

At the outset, our problem appears impossible. We do not even know whether the two lines intersect in space! Even if we somehow knew that they did, we still don't know how to go from the proximal angle φ to the distal angle θ. Clearly there can be no deterministic rule that takes us from proximal to distal angles, for there are many values of distal angle θ that all lead to the same value of the proximal angle φ. But perhaps we can, given the proximal angle φ, rank the possible values of the distal angle θ as candidates by some measure of plausibility.

To do so, we need to understand how the scene was generated. Let's imagine that the scene is completely under the control of a "Demon" who first draws the angle on a transparent sheet of acrylic (Fig. 1.2), perpendicular to the line of sight of the observer, and then gleefully spins the sheet of acrylic about a horizontal axis. The spin is so forceful that we can assume that the acrylic sheet ends up at any slant angle σ as likely as any other. While the angle is drawn and the sheet spun, the observer waits patiently with eyes closed. The angle that the Demon drew is symmetric around the vertical axis before and after the spin as illustrated in Fig. 1.2. At the Demon's command, the observer opens his or her eyes and, challenged by the Demon, must specify the true angle θ given that only the projection φ is seen.

Given this generating model, we can say quite a lot about the relation between φ and the possible θs that might have given rise to it. As the acrylic sheet gets more slanted relative to the line-of-sight, the projected angle φ in Fig. 1.2 increases to 180 degrees (when the sheet is horizontal) and then decreases back to the true value, when the acrylic sheet is upside down. The projected angle φ is always greater than or equal to the true angle θ. We can say more than this, though. Given the known value of φ and what we know about the generating model, we can compute the probability that any particular value of the unknown θ was the angle that the Demon drew. This probability distribution is the first element of Bayesian modeling that we introduce and it is the *likelihood function.*

Likelihood function

Let us assume that the angle φ between the two lines in the image is measured to be 135 degrees. What is the likelihood that any given world angle θ, say, 95 degrees, was the angle that the Demon chose to draw? To answer this question, we can imagine the following simulation. We take our own acrylic sheet with two lines painted on it separated by 95 degrees and we then spin it ourselves, repeatedly, simulating what the Demon did, over and over. Following each simulation of the Demon's spin, we measure the resulting angle φ. If the projected angle is approximately[1] the angle φ that we measured (say, 135 degrees plus or minus 5 degrees), we decide that the simulation is a "hit", otherwise it is a "miss". Repeating this procedure 1000 times, we end up with 5 hits and 995 misses. We conclude that the probability of obtaining an image angle of approximately 135 degrees given a world angle of 95 degrees is about 0.005.

If we repeat our simulation experiment for all possible values of the world angle, we obtain a histogram estimation of the *likelihood function* (Fig. 1.3). The likelihood is a function of both the known image angle φ and the unknown world angle θ.

Note that θ is a random variable whereas φ is considered as a fixed parameter (cf. [1]). For this reason, we prefer to think of the likelihood function as a function of the distal angle θ indexed by the proximal angle φ. It is important to understand that the likelihood function is not a probability distribution function since the integral over all values of the distal angle θ need not be 1; it is the integral over all values of the proximal angle φ that must equal 1. We introduce the following notation for the likelihood function:

$$\text{likelihood}_{\varphi}(\theta) = p(\varphi \mid \theta). \tag{1.1}$$

1. Note that angle is a continuous variable, so we are dealing with a continuous distribution. The probability of any particular value (e.g. an image angle of precisely 135 degrees) is zero. Such distribution functions are not easily estimated using Monte Carlo simulations without approximating them as discrete distributions (e.g. binning the angles into bins of width 10 degrees).

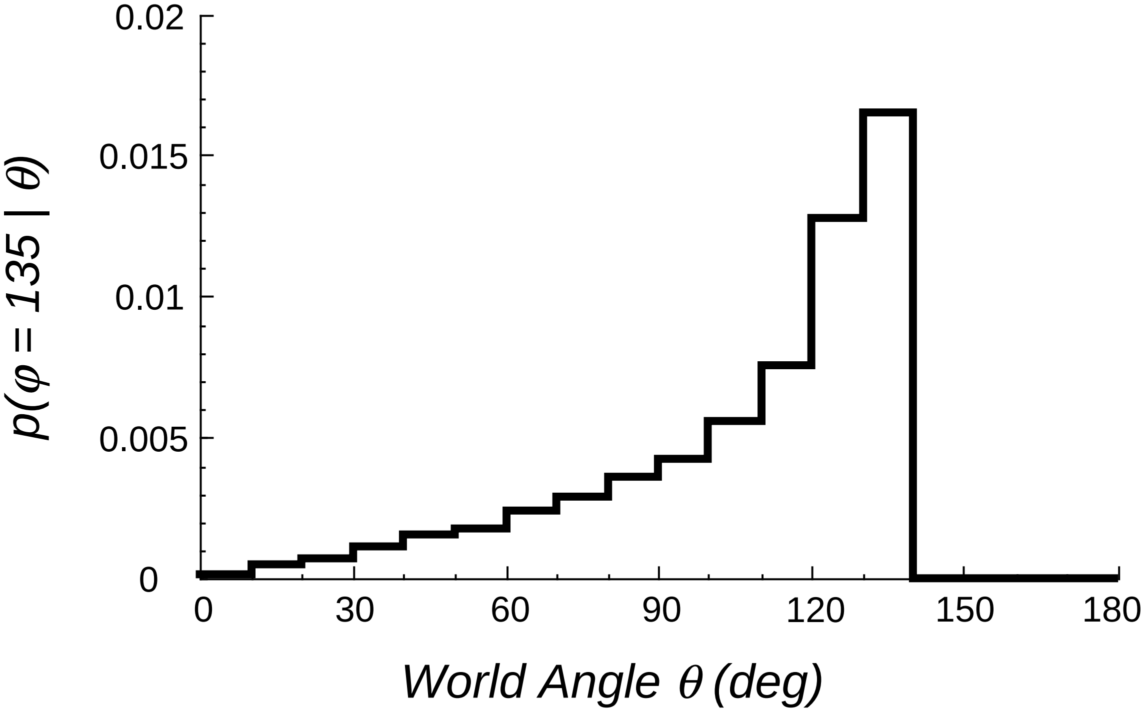

Figure 1.3: The likelihood function (when the image angle is 135 degrees). The likelihood characterizes the chances that a particular world angle projects as a given image angle.

The Demon is still waiting for a response. Given the likelihood function in Fig. 1.3, it would be foolish to guess that the Demon drew an angle as small as, say, 10 degrees. The likelihood that a world angle of 10 degrees would lead to the observed value of φ (135 degrees) is very small relative to larger angles. The likelihood increases for world angles up to 135 degrees at which point the likelihood drops to zero. (Recall that the reason for the zero values above 135 degrees is that the projected angle is always larger than the world angle so it is impossible for a world angle larger than 135 degrees to produce an image angle of 135 degrees. The largest likelihood is therefore reached for a world angle equal to 135 degrees.) The world angle that is most likely to have given rise to the observed φ is 135 degrees.

If we choose this angle as our estimate of the response variable, we follow the strategy called Maximum Likelihood Estimation (MLE) [17]. We choose, as our response, the world angle that had the highest probability of generating the observed sensory information. This decision rule is close in spirit to the Likelihood Principle proposed by Helmholtz [4].

The likelihood function captures all of the sensory information we have. What happens if we have extra knowledge or beliefs beyond the simple geometrical knowledge included in the likelihood function? Suppose that we know that our Demon has a marked bias toward world angles near 90 degrees and an aversion toward smaller or larger angles.

Prior distributions

We represent our knowledge about the Demon's preference for some world angles over others as a *prior distribution*, a probability distribution function on the response variable:

$$\text{prior}(\theta) = p(\theta). \tag{1.2}$$

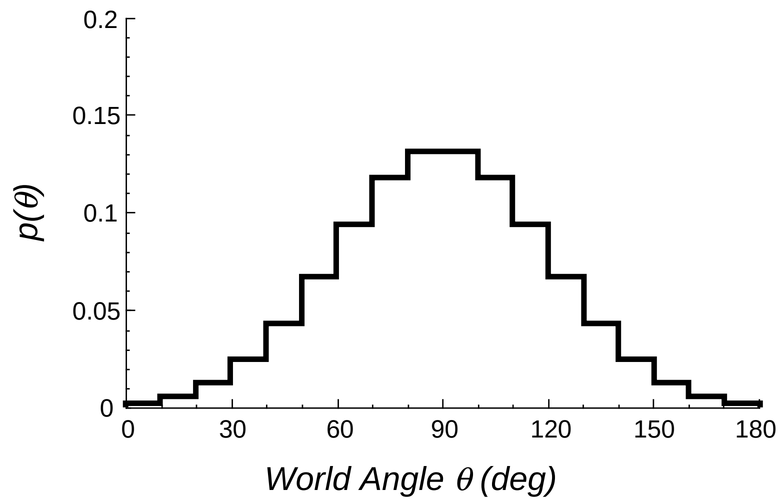

Figure 1.4: Prior distribution (discretized Normal distribution with a mean of 90 deg and a standard deviation of 30 deg). The prior distribution represents the prior belief of the observer as the frequency of occurrence of various world angles.

Figure 1.4 shows one such prior distribution chosen to be a Gaussian distribution with a mean of 90 degrees and a standard deviation of 30 degrees. The value of the standard deviation in effect represents how strong the Demon's preference for 90 is. When the standard deviation is very small, the resulting Gaussian will be almost a spike at 90. If this distribution accurately captured the Demon's behavior, then we would know the Demon almost always picked an angle of 90 degrees or very close to it. Even without sensory information, we would expect to be able to make a very good estimate of the world angle. If, on the other hand, the standard deviation is very large, the resulting Gaussian tends to a uniform distribution. This case represents a very slight preference of the Demon for 90 degrees such that knowledge of the prior is not helping us much in our estimation task. We shall assume for now that the Demon's preference is well captured by a Gaussian with standard deviation of 30 degrees as shown in Fig. 1.4: the bias toward 90 is strong but not so strong as to tempt us to ignore the sensory information embodied in the likelihood function.

So far, we have encountered two elements found in any Bayesian model: the likelihood function and the prior. The former represents everything we know about the process that turns the state of the world into sensory information. The latter describes what we know about the relative plausibility of different states of the world. Both represent knowledge in probabilistic terms, permitting us to combine sensory information and prior knowledge about the world using probability calculus.

Bayes' theorem and the posterior distribution

The Demon is still waiting for an answer. We next need to combine likelihood and prior to compute a posterior distribution, our estimate of the probability of different states of the world (here, θ) given everything we know. The following formula describes how to compute the posterior distribution:

$$\mathrm{posterior}_{\varphi}(\theta) = C \times \mathrm{likelihood}_{\varphi}(\theta) \times \mathrm{prior}(\theta), \tag{1.3}$$

where C is a constant chosen so that the posterior is indeed a probability distribution function (i.e., the integral over all possible values of θ is one). Intuitively, the posterior of one world angle θ is simply the likelihood weighted by the prior probability of occurrence of θ. The formula is derived

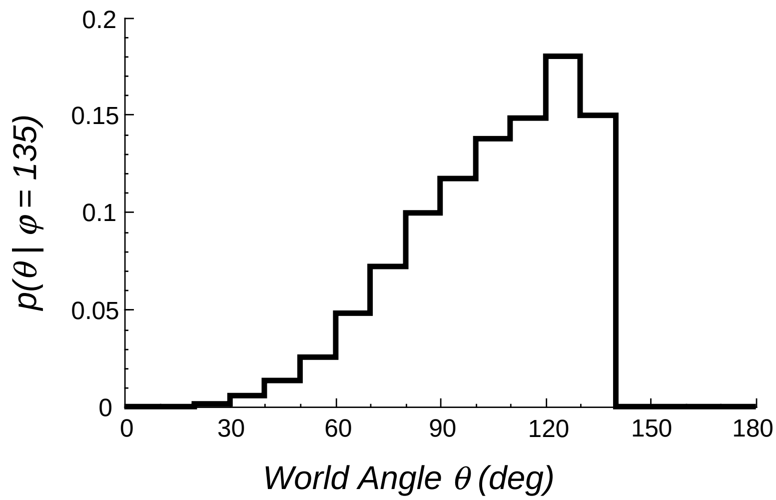

Figure 1.5: Posterior distribution (given that the image angle was 135 degrees). The posterior distribution represents the likelihood of occurrence of the response variable weighted by prior beliefs.

from Bayes' Theorem, a simple mathematical result concerning computations with conditional probabilities:

$$p(\theta \mid \varphi) = p(\varphi \mid \theta) \times p(\theta)/p(\varphi). \quad (1.4)$$

The left-hand side is the posterior, $p(\varphi \mid \theta)$ is the likelihood, and $p(\theta)$ is the prior.

Comparing the last two equations, we note that C equals the inverse of $p(\varphi)$, but it is easy to avoid computing it (or even thinking about it) since (1) it is the constant needed to make the posterior distribution a distribution, and (2) given the uses to which the posterior is put, it often turns out that the normalization by C is not necessary.

The posterior distribution is our estimate of the probability distribution of the unknown angle θ after both the sensory data and the prior distribution are taken into account. Fig. 1.5 shows the posterior distribution when the likelihood function was computed as in Fig. 1.3 and the prior probability was chosen as in Fig. 1.4 . Note that the peak of the posterior distribution is located between the peaks of the likelihood function (Fig. 1.3) and the prior distribution (Fig. 1.4).

In effect, the two sources of information we have are being pooled. The likelihood function would lead us to pick values of θ near 135 degrees as our maximum likelihood guess, the prior would lead us to favor a guess nearer 90 degrees, reflecting knowledge of the Demon's preference. The two pieces of information average out to a posterior whose peak falls around 125 degrees, about 22% of the way from the likelihood peak to the prior peak.

We could imagine redoing this example with different values of the standard deviation of the prior. When the standard deviation is smaller, the peak of the posterior moves toward the peak of the prior (i.e., 90 degrees), when it is larger, the peak of the posterior moves toward the peak of the likelihood function (i.e., 135 degrees). The standard deviation of the prior is effectively controlling the relative importance of prior and sensory information. When the Demon's preference toward 90 is known to be very strong, we should favor our prior knowledge and downweight conflicting sensory data. Alternatively, we should favor the sensory data when the prior is weak.

Intelligent selection of the distributional form for the prior is an important part of modeling. Within the Bayesian framework, the prior encodes all we know about the state of the world independently of the current sensory input.

Still, the Demon is becoming impatient. The posterior distribution is what we have, but what we need is an estimate of θ, a "best guess". That's what the Demon wants from us.

Gain, loss, and the decision rule

The decision rule links the posterior distribution with the action taken by the observer. The action can be an explicit motor act, such as orienting the hand appropriately to grasp an object, or an estimate of some aspect of the world such as the angle θ. One possible criterion for choosing an action is to pick the mode of the posterior distribution, the most probable value of θ according to the posterior distribution. If we used such a *Maximum a Posteriori (MAP) Rule*, our response to the Demon's challenge would be "125 degrees". This rule is widely used in the Bayesian literature and it is a plausible choice of decision rule.

Within the framework of Bayesian Decision Theory the choice of decision rule is remarkably simple and principled. It begins by assuming that, for every action we might take, there are consequences that depend upon the true state of the World: a numerical gain or loss, where a loss is merely a negative gain. In the example we have been pursuing, we have neglected to describe the consequences of our actions. Let us remedy this deficiency now.

Suppose that, if we guess the Demon's angle θ to within, say, 5 degrees, he goes away ($+100$, a gain). If we miss by more than 5 degrees, he draws a new angle, spins the acrylic sheet anew, and continues to torment us (-50, a loss). We shall call such a specification of gain and loss a *gain function*. We can then choose the action that maximizes our expected gain (cf. also chapter 6 by Nadal). The *Bayes decision rule* is precisely the choice of action α that, given the posterior, maximizes expected gain:

$$\text{expected\_gain}_{\varphi}(\alpha) = \int_{\theta} \text{posterior}_{\varphi}(\theta) \times \text{gain}(\alpha, \theta) d\theta. \tag{1.5}$$

The particular gain function we chose in this example is an example of a *discrete Dirac* or *Delta* gain function which takes on two values only. The larger gain is obtained when the angle estimate (the action) is within a small interval surrounding the correct angle, otherwise the lesser gain is obtained. If the small interval is infinitesimally small, then, it turns out that the Bayes decision rule corresponding to the Delta gain function is precisely the MAP rule. When the interval is small but finite, as in our example, the Bayes decision rule is approximately the MAP rule. Thus, if we wish to be rid of the Demon as quickly as possible, we should pick the most probable value of θ according to the posterior distribution. Any other choice prolongs, on average, our encounter with the Demon.

But suppose that the gains and losses are different from those just specified. What if the Demon appears with his acrylic sheet and announces that, today, he will force the observer to pay him, in dollars, the square of the difference in degrees between his estimate and the true value of θ. A five degree error costs $25, a 10 degree error, $100. This gain/loss function is the *least-square loss function*, a very commonly used loss criterion in statistical estimation. The Observer's immediate question is likely to be "Is the MAP still the best rule to use?" The answer turns out to be no. The rule which minimizes his expected loss now is to pick the *mean* of the posterior distribution, not the mode. The Bayes rule associated with the least-square loss function is the mean. The mean of the posterior distribution in Fig. 1.5 is 105 degrees, significantly less than the MAP estimate, which was 125.

We have already noted that the prior distribution is responsible for the shift of the peak of the posterior probability away from the peak of the likelihood function. If the prior distribution is uniform (i.e. there is no evidence for any bias for the Demon's behavior), the maximum of the posterior equals the maximum of the likelihood. In this limit case, the maximum a posteriori (MAP) rule is identical to the maximum likelihood estimate (MLE) rule.

While the Delta gain function and the associated MAP rule are appealing for their simplicity, this decision rule is sometimes prone to favor unstable interpretations such as scenes viewed from accidental viewpoints [3, 18]. One possible cure to this problem is to choose a loss function that gracefully increases when the estimate of the world angle α deviates from the real world angle θ. The least-square loss function described above is one such function.

It is important to emphasize that there is not one computation applied to the posterior distribution that defines uniquely the Bayes rule. Rather, the rule varies as the gains and losses vary. The Bayes rule is always: *do whatever you must to maximize expected gain*.

The Bayesian framework allows us to model the consequences of gains and losses on behavior and to represent performance in different tasks in a common framework so long as the effect of a change of task can be modeled as a change in possible gains and losses. See Berger [1] or Maloney [9] for a more detailed discussion.

One important property of the decision rules we have discussed so far is that the same action will be chosen whenever the same stimulus φ is seen. This must be contrasted with the variable behavior of any biological organism. There are at least two approaches to modelling this variability. The first approach is to recognize that we have neglected to model photon noise and sources of noise in the visual system. Introduction of these as contributors to the likelihood component of a Bayesian model would introduce variability into human responses to identical stimuli. The second approach is to abandon the Bayes rule as a realistic decision rule. For instance, Mamassian & Landy [11] have modeled human response variability with what they termed a "non-committing" rule. According to this non-deterministic rule, an action is chosen with a probability that matches its posterior probability. Actions with high posterior probabilities are selected more often than those with low posterior probabilities, but any action may potentially be chosen. This decision is also known as *probability matching* [13] and is often observed in human and animal choice behavior when the gain function is the Delta gain function described above. It is important to note that this rule is not optimal (the MAP rule always leads to higher gain). Even though the MAP Rule is optimal in this case, humans and other animals persist in probability matching to a remarkable degree. In fact, it might be a better strategy for an animal since it allows exploration of the state space for learning (for an introduction to learning by exploration, see [16]).

Discussion

In this section, we have defined the three fundamental steps in Bayesian modelling of visual perception (Fig. 1.6). The first step is to define the prior distributions to represent what we know about the probability of encountering different states of the world. The second step is to compute the likelihood function by determining all the ways the stimuli could have been generated and their frequency of occurrence. Using Bayes' theorem, likelihood and priors can be combined to produce a posterior distribution, i.e., the best estimate of the probability of each possible state of the world given the sensory data and the prior distribution. The third and last step is to settle on a decision rule that transforms the posterior distribution into an action. In a full Bayesian model, this decision rule is computed from Bayes' rule and is fully described by the gain function.

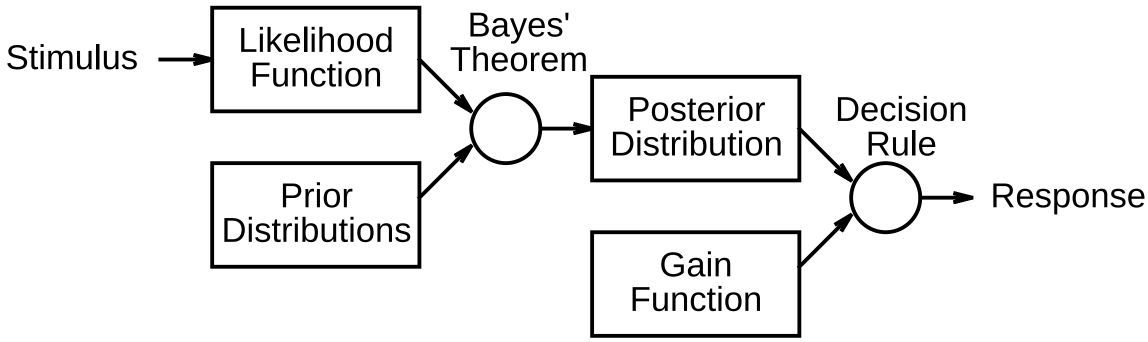

Figure 1.6: Flowchart of the steps involved in Bayesian modelling.

The Study of Human Priors: Perception of Embossed Surfaces

Issues

In the previous part of the chapter, we have provided a model of perceptual decision making based on sensory information, *a priori* knowledge, and the choice of a decision rule. The critical question is, of course, "Are Bayesian models appropriate for human vision?" While it is too early to answer this question conclusively, a first test of the model is to analyze the compatibility of the model with human data. At this point, it is important to keep in mind that even if visual processing could be perfectly represented as a Bayesian model, the modeler's choice of distribution, likelihood, gain function, or decision rule may be different from what is actually embodied in visual processing. Our first endeavor should therefore be to learn more about the form of the prior distributions, gain functions, and decision rules that accurately model human vision.

Assuming for now that we know how to compute the likelihood function and the decision rule, we shall use Bayesian models to investigate how prior knowledge is encoded in the visual system. As a first guess, it is convenient to model prior knowledge as a Gaussian distribution (or von Mises distribution for periodic variables such as angles; cf. [12]). The advantage of the Gaussian distribution is that it is fully parameterized by only two parameters, namely its *mean* and *variance*. This prior encodes a preference in favor of the mean value and values near it. The variance of the Gaussian is inversely related to the strength of the evidence in favor of the prior mean. As the variance increases, the prior distribution converges to the uniform distribution. It is important to note that while the Gaussian distribution is everywhere non-zero, other non-Gaussian priors can be zero across an interval; in this case, no amount of sensory data can beat the prior.

In this section we consider simple visual tasks that are sufficiently ambiguous that we can hope to observe the influence of prior distributions in visual performance. We summarize here our work on the perception of shaded embossed surfaces with parallel contours [10, 11, 12].

Assumption of illumination from above-left

Light sources are usually located above our head. It is commonly believed that this regularity of the illumination position is at the origin of the change in perceived relief of figures that are turned upside-down [2, 14]. We have been interested in quantifying this prior assumption used by the visual system [10]. We report the results of this experiment with reference to the Bayesian framework developed in the previous part of this chapter.

Stimuli were shaded embossed surfaces as shown in Fig. 1.7. These stimuli can be interpreted as

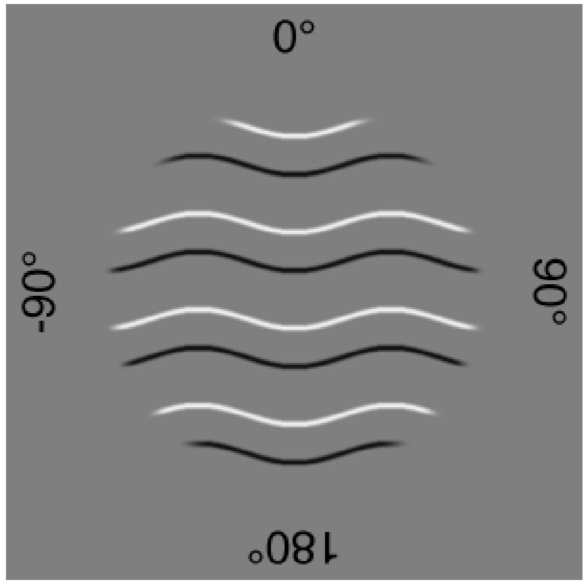

Figure 1.7: Stimulus used in the illumination experiment. This figure was presented at different orientations in the image plane. Observers indicated whether they perceived the wide or narrow strips bulging in relief.

planar frontal surfaces with either narrow or wide strips bulging towards the observer. When the figure is rotated in the image plane, the perception alternates between these two interpretations. The fact that the very same figure can be interpreted in two different ways with a mere rotation indicates that the visual system is using additional knowledge beyond the stimulus information. We presented the figure at different orientations to human adult observers and asked them whether they perceived the bulging strips as being narrow or wide. Each orientation was presented multiple times to estimate the variability of responses for each orientation.

Results for one observer are shown in Fig. 1.8. Let us call *narrow score* the proportion of times the stimulus was interpreted as formed by narrow strips bulging. The plot shows the narrow score as a function of the orientation of the figure in the image plane. It is clear from the plot that responses depended strongly on the orientation of the stimulus. When the bright contours of the figure are interpreted as edges facing the light source and dark contours in shadow, the peak narrow score should occur when the stimulus orientation is zero. Remarkably, the center of the peak of the narrow score is shifted to the right of zero, corresponding to a stimulus most consistent with light coming from above and slightly to the left of the observer (for a similar demonstration of an above-left illumination bias, see [15]).

The smooth curve in Fig. 1.8 shows the best fit of the Bayesian model to the human data. The model has knowledge about the shape of the object, the illumination and viewing conditions. The object is modeled as a flat surface with strips in relief. The narrow strips are either raised or lowered relative to the background flat surface and the orientation of the edge between the strip and the background defines the *bevel angle* (positive values correspond to the narrow strips being in relief). The illumination conditions are modeled with an ambient and a point light source of different intensities. In addition, the illumination model includes commonly used shading assumptions (uniform Lambertian reflectance). Finally, the viewpoint is modeled as the orientation of the viewing direction relative to the surface, thereby disregarding the distance between viewer and object.

The object, illumination, and viewing models have degrees of freedom that can be described as

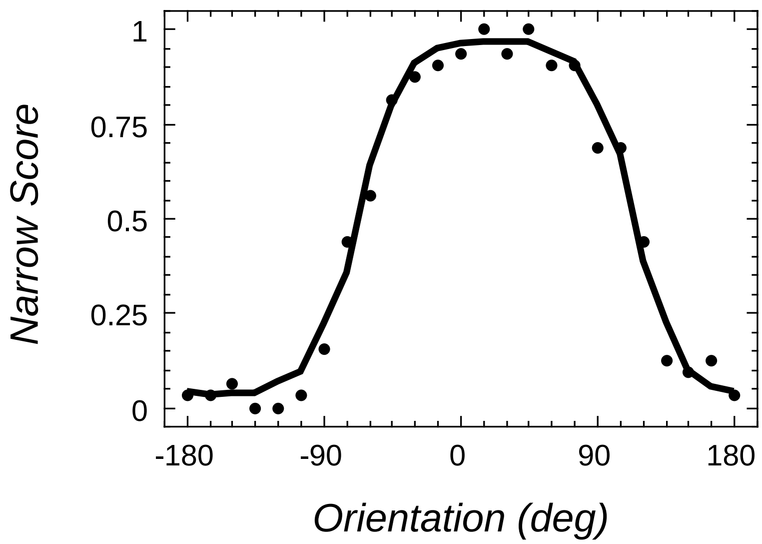

Figure 1.8: Results for the illumination experiment (for one observer). The dots indicate the proportion of times an observer perceived the narrow strips in relief, and the solid line shows the performance of the model.

free parameters. Each parameter has its own prior distribution. For most parameters, we have no *a priori* knowledge and hence assume uniform prior distributions for these parameters. To model the prior assumptions that we anticipated to exist, we gave the corresponding parameter a non-uniform distribution, generally a Gaussian distribution specified by a mean (the value toward which observers were biased) and a variance (inversely related to the effectiveness). In this way, we allowed for biases on the thickness of the strips in relief (i.e., bevel angle), the light source tilt (i.e., light direction from approximately above) and the surface tilt (i.e., view-from-above, discussed in the next section).

The model calculates the posterior probability $p(narrow \mid stimulus)$ by first calculating that probability for each possible combination of illumination and viewpoint parameters, and then summing them:

$$\int p(narrow, illumination, viewpoint \mid stimulus)\, d(illumination)\, d(viewpoint). \quad (1.6)$$

The integrand is expanded using Bayes' theorem into a likelihood and a prior term. The prior probability term, $p(narrow, illumination, viewpoint)$, is further expanded by assuming that illumination and viewpoint parameters are independent. Finally, the output of the model is obtained after application of the non-committing decision rule. That is, the proportion of times observers respond "narrow" is the same as the calculated posterior probability that the stimulus arose from an object with bulging narrow strips.

Each parameter can now be estimated from the best fit of the model to the human data. The advantage of the model is that each parameter has a straightforward meaning. The best-fitted parameter for the light source direction indicates a bias of approximately 25 degrees to the left of the vertical. The effectiveness of this bias on the light source position and the bias for the viewpoint can be estimated by looking at the variance of the corresponding prior distributions. In addition, there was a very slight bias to prefer narrow rather than wide strips in relief.

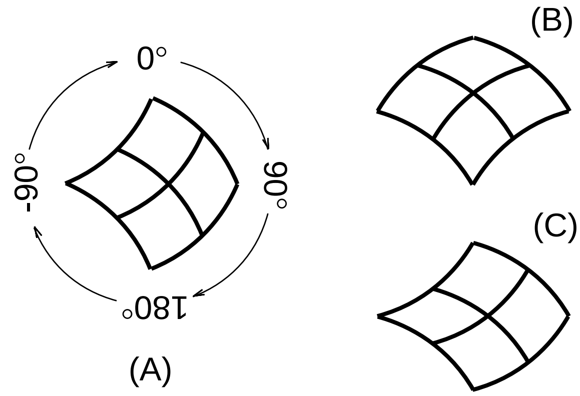

Figure 1.9: Stimuli used in the viewpoint experiment. The line drawing shown in A can be interpreted as an egg-shaped patch or saddle-shaped patch depending on its orientation in the frontal plane. Figure B tends to be seen as egg-shaped more often, whereas C is more often seen as saddle-shaped. Each stimulus was presented at different orientations in the image plane. Observers indicated whether they perceived an egg-shaped or saddle-shaped patch.

Assumption of viewpoint from above

Using the line drawing shown in Fig. 1.9, we showed that observers have a bias to interpret contour drawings as if they were viewing the object from above [11]. In this experiment, observers reported whether the line drawing appeared to be the projection of a surface patch which was egg-shaped or saddle-shaped. Again, these images were shown at various orientations in the image plane, and observers indicated whether they initially perceived the object as egg- or saddle-shaped. We rely on the initial interpretation, averaged across multiple presentations of the same stimulus, to determine the observer's propensity to perceive the stimulus in one way or the other.

The data in Fig. 1.10 clearly showed that responses again depended strongly on the stimulus orientation. The preference for a viewpoint located above the scene can be modeled as a preference for the normal to the surface to point upwards [11]. This preference can therefore be modeled as a bias on the tilt of the surface normal, with the preferred tilt equal to 90 degrees. The bias on the surface normal results in a bias to interpret contours that are convex-upward in the image as being convex toward the observer.

We again fit a Bayesian model using exactly the same strategy as in the illumination case above. Here, the scene description included parameters related to the surface shape (same-sign principal curvatures correspond to egg-shaped surfaces), the viewpoint (i.e. the surface orientation as defined by its slant, tilt and roll), and the way surface contours were painted on the surface patch (defined by their orientation relative to the principal lines of curvature). Most parameters were given uniform prior distributions except for the surface tilt (corresponding to our suspected bias for a viewpoint above the object), the surface curvature (for a good fit we required a bias for perceiving 3D contours as convex) and the surface contour orientations (which were biased to be closely aligned with the principal lines of curvature). Again, the non-committing rule was used. From the best fit of a Bayesian model that included these prior constraints, we estimated the standard deviation of the prior distribution on surface tilt orientation to be approximately 30 degrees.

The prior assumption on the surface tilt probably plays a role in the interpretation of the Necker cube shown in Fig. 1.1A. We have already noted that only two interpretations out of an infinite number are preferred by human observers (Figs. 1.1B and 1.1C). With prolonged viewing, observers alternate between these two interpretations with a frequency that can be well-modeled

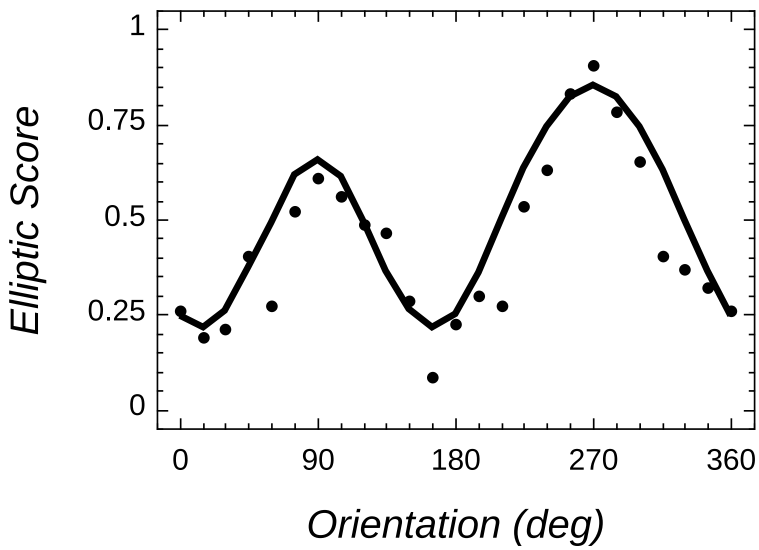

Figure 1.10: Results for the viewpoint experiment (averaged over 7 observers). The elliptic score (i.e. the proportion of times observers indicated Fig. 1.9A was egg-shaped) is shown as a function of the orientation in the image plane. The best fit of the Bayesian model is shown as a solid line.

by a Gamma distribution. The mean time spent in each interpretation, however, is different. One direct prediction from the results of this section is that the preferred cube interpretation will be the one for which most of the normals to its faces point upwards.

Interaction of prior assumptions

In the last two sections, we have seen that the interpretation of ambiguous figures is guided by two assumptions: a preference for the illumination to be located above (and to the left) of the observer and a similar preference to have the viewpoint located above the scene. In this section, we look at stimuli for which both of these assumptions can be used to disambiguate the figure. For some orientations of each stimulus the two prior constraints were in agreement as to the preferred interpretation of the stimulus, and in others they were in conflict. An example stimulus is shown in Fig. 1.11 [12].

These stimuli can again be perceived as embossed surfaces with either narrow or wide strips in relief. The data show that the observer's interpretation changed as the figure was rotated in the frontal plane (Fig. 1.12).

The human performance was well modeled by a Bayesian model that included both the illumination and the viewpoint priors described in the previous sections. The model was of exactly the same form as the previous two, including parameters for the surface shape (i.e. the bevel angle), the illumination and viewpoint. Again, the non-committing rule was used and the parameters of the prior distributions were varied to best fit the data. The best fit of the model allowed us to extract the parameters of the prior distributions. The light direction bias was 10 degrees to the left of the vertical (the viewpoint bias was set to 90 degrees for the surface tilt). The standard deviation of the illumination prior was about 40 degrees, whereas the standard deviation of the viewpoint prior was about 80 degrees. In addition, there was a very slight bias to prefer narrow over wide strips in relief.

In the first part of this chapter, we argued that the standard deviation of a prior distribution is inversely related to the effectiveness of this prior. We tested whether this effectiveness could be

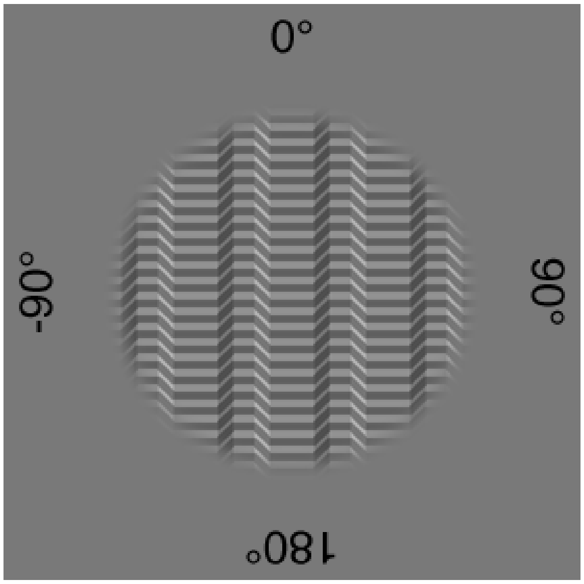

Figure 1.11: Stimuli used for the interaction of priors experiment. Depending on its orientation in the frontal plane, this figure can be interpreted as an embossed surface with narrow or wide strips in relief. Both shading and contour cues to shape are present in the stimulus.

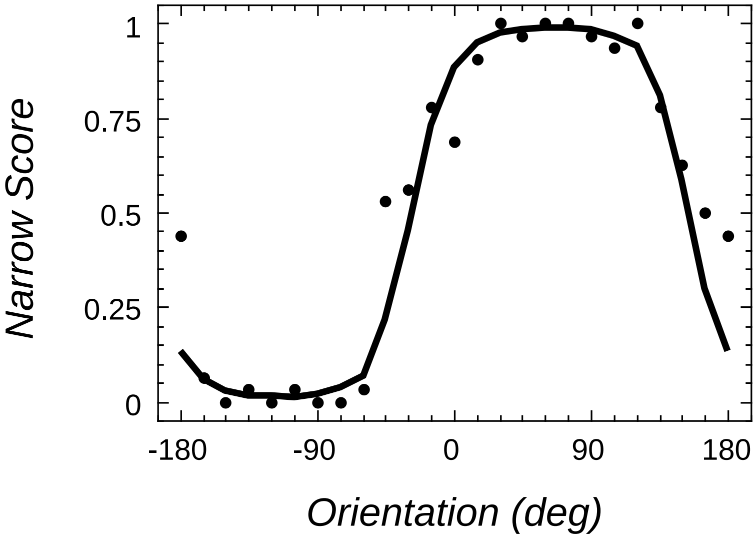

Figure 1.12: Results for the interaction of priors experiment (averaged over 8 observers). The figure gives the proportion of trials observers indicated that the narrow strips were seen bulging as a function of the image plane orientation of the stimulus. The solid line shows the best fit of the Bayesian model.

affected by changing properties of the stimulus. We repeated the prior interaction experiment for stimuli for which the stimulus contrasts of the shading and surface contours were independently varied. We fit each contrast condition separately. We found that increasing the shading contrast decreased the standard deviation of the illumination prior, and similarly, increasing the contour contrast decreased the standard deviation of the viewpoint prior. These results are consistent with the notion that the effectiveness of a prior distribution reflects the reliability of the particular cue with which it is associated (see chapter 3 by Jacobs for a related finding). It is important to realize that we have treated these prior distributions as variable things, dependent on aspects of the stimulus itself. The apparent cue reliability that drives this change can be estimated from ancillary measures extracted from the stimulus (e.g., the contrast of the shading or contour edges in our example).

Discussion

In this part of the chapter, we have summarized three studies that looked at the 3D perception of ambiguous images. We used the Bayesian framework developed in Section 1 to focus on the prior assumptions used by the visual system. We found a preference for the illumination to come from above-left and for the viewpoint to be located above the scene. The Bayesian models allowed us to explain human performance quantitatively.

We also looked at a situation in which the consequences of two prior assumptions must be combined. We found that our framework was still applicable in predicting human performance, but only if we allowed the standard deviation of the prior distributions to be affected by the stimulus properties. In this experiment, different parts of the stimuli changed contrast but the task of the observer was kept constant. Therefore, we did not change the gain function or the decision rule to try to explain the shifts of human criteria with stimulus contrast. In addition, while the likelihood changed with the stimulus contrast, this change was not large enough to explain the criteria shifts. Only a change in the prior distribution could reasonably explain the changes in human performance. More specifically, we found that the prior confidence value as measured by the variance of the prior distribution was inversely related to the contrast of the stimulus attributes relevant for this prior. This finding is important because it forces us to reconsider a basic tenet of Bayesian modelling that prior constraints are pieces of knowledge that are independent of the stimuli and of the task with which observers are confronted. This issue is further discussed in the next section.

General Discussion

In this chapter, we have described a general framework to study the resolution of ambiguities arising in three-dimensional perception. We have shown how the framework can be implemented by describing a few applications. The framework is Bayesian in the sense that it emphasizes the role played by prior assumptions. In this last part of the chapter, we discuss some of the issues that are inherent in this framework.

Is this really Bayesian modeling?

The models we discussed in Section 1 were cited as examples of Bayesian modeling of depth disambiguation. In fact, they stray fairly far from a strict Bayesian interpretation of a decision making problem. We now describe the Bayesian and non-Bayesian components of our models, and discuss our motivation for departing from the strict Bayesian approach.

Our models begin in the Bayesian spirit by defining prior distributions on various parameters that give rise to the stimulus situation. These parameters describe the objects of the physical environment, the illumination conditions and the viewing geometry. More specifically in our examples, objects are parameterized by the bevel angle, reflectances of the various surfaces, principal surface curvatures, orientation of the surface contours relative to those surface curvatures, and surface shape. The illuminant is simply characterized by the position of the light sources and their intensities. Finally, the viewing geometry is parameterized as the orientation of the observed object relative to the viewer (the viewing distance should also be included in perspective projection). When we have no reason to expect observers to have a particular bias, we give the corresponding parameter a uniform prior distribution. In contrast, where biases are expected, the prior is modeled as a non-uniform distribution where the modes of the distribution match the biases. Typically, we anticipate a single bias along each dimension, so that the prior distribution is unimodal (e.g. Gaussian). For instance, among all possible light source directions in the frontal plane, only one direction is preferred. Moreover, we showed how these biases can be estimated experimentally [11, 12]. So far, this is pretty much the Bayesian methodology.

However, we depart from a strict Bayesian methodology in a number of ways. Let us first consider the likelihood function. The likelihood characterizes the mapping from world to retinal image. When the stimulus has very low contrast, photon statistics will affect the image registered on the retina. Moreover, internal noise will also affect the way stimuli are represented at successive stages in the visual system (see chapters 2 by Schrater & Kersten and 4 by Weiss & Fleet). These external and internal sources of noise will cause the proximal stimulus to be stochastic and as a result, the computed likelihood will vary from trial to trial, and the observer's decision will vary when the same distal stimulus is repeatedly presented. In contrast, our likelihood function is deterministic. Either this set of object, illumination and viewing parameters gives rise to this image or it does not. We argued that with our high-contrast stimuli, external noise is negligible and internal noise alone can not reasonably explain the large variations in subjects' responses.

Our second departure from a strict Bayesian approach is the way we treat the decision rule. In a sense, this follows directly from our choice of a deterministic likelihood function. If the decision rule is a Bayes rule, it is also deterministic. Examples of Bayes rules include the maximum likelihood and the MAP rules. When the likelihood and decision rules are both deterministic, the observer will give the same response to trials that use the same stimuli. This prediction was clearly violated in our psychophysical experiments. Therefore, instead of following a Bayes rule, we chose a stochastic rule that we termed a "non-committing decision rule". According to this rule, the observer first computes the posterior probability of each possible scene given the image and then chooses each response with probability equal to its estimated posterior probability. If, for instance, there are only two possible scenes whose posterior probabilities are 0.7 and 0.3 respectively, then the observer will choose the first scene with a probability equal to 0.7. This rule is suboptimal (unless the gains are such that incorrect answers are not penalized) and hence is not a Bayes rule. However, it is a behavior that is commonly found in humans and animals, and is referred to as *probability matching* [13].

Finally, we depart from a Bayesian approach by the way we treat prior constraints. We looked at the effect of context on prior constraints by varying the stimulus contrast along different parts of an image (Section 1). We found that the effect of the prior associated with the shading cue to depth increased with shading contrast (and similarly for the contour cue to depth). We fit the model to the data separately for each contrast condition, allowing the parameters for the priors (light-from-above and viewpoint-from-above) to vary between fits. That is, we allowed the prior distributions to depend on the stimulus. This is distinctly *not* a Bayesian operation. If the priors depend on the stimulus, then they are not "prior"! Nevertheless, the results of this exercise were quite interesting. The variance of the prior distributions, as we have discussed, is inversely related to their effectiveness in biasing the overall estimate. We found that this effectiveness was a monotonic function of contrast. This makes sense, even though it is entirely contrary to the notion that these were parameters of prior distributions.

In summary, we depart from the traditional Bayesian approach by the way we treat the three basic components of any Bayesian model. We believe that the reasons for these departures were well-motivated for the experiments presented in this chapter. Future research will tell us whether our choices are valuable in other contexts.

Future directions

In this chapter, we have described Bayesian-inspired models of visual perception. These models provide an explicit account of the interaction between sensory information and prior assumptions, as well as the decision rule used by the observer in a given task. Each component of the models is justifiable and experimentally testable. Therefore, Bayesian models appear psychologically more relevant than previous models that relied on *ad-hoc* components such as the "smoothness constraint" (in fact, other frameworks such as regularization are just special cases of the Bayesian framework; [18]). In spite of its great appeal, Bayesian modelling is still in its infancy. We now discuss some future directions of research.

The first issue deals with the interaction between the likelihood function and the prior distributions. In this chapter, we have chosen a likelihood function that faithfully takes all the information in the image and that is not affected by noise. This choice implies a highly peaked likelihood function and a moderate influence of the prior distributions on the posterior distribution. Alternatively, the same posterior distribution could have been obtained by a more shallow likelihood function and a greater influence of the prior distributions (cf. the influence of external noise on the likelihood function in [7]). More work is needed to understand better how different contextual situations (such as different amounts of external noise) affect the likelihood function and the prior distributions (cf. chapter 7 by Yuille & Coughlan).

Another issue deals with the origin and stability of prior constraints. Do prior constraints directly reflect the statistical regularities of the environment? If so, we can foresee that priors will be conditional on context (e.g. a forest prior vs. a city prior). When more than one prior can be applied in a certain context, then these priors will have to compete (for an example on shape from shading, cf. [18]). Another related question is whether priors can be updated when the environmental conditions change. Updating a prior constraint means that the visual system has the ability to sample and memorize an extremely large number of scenes (cf. [19]). This strategy is impractical unless some sort of approximation is used to build the new priors. Finding biologically plausible implementations of Bayesian models and their approximations is an area of active research (see the chapters in the second part of this book).

One final important issue deals with the importance of the task with which the visual system is faced. It is obvious that different tasks correspond to different decision rules. Some motor tasks might put the organism more at risk than purely visual exploration tasks, thereby affecting the cost of making the wrong decision. However, it is perhaps less obvious that the task could also affect the prior constraints used. Different tasks might indeed direct the attention of the visual system to different components of the stimulus and their associated priors. Future research should look at the level(s) within the visual system at which the choice of task has influence.

To conclude, we have reviewed the main principles underlying Bayesian modelling of three-dimensional perception. The framework entails the derivation of the likelihood function, the description of prior assumptions and the choice of a decision rule. We believe that this framework provides a powerful tool to study the mechanisms involved in human visual perception and that it can have important generalizations in others areas of cognitive neuroscience.

Acknowledgments

Funding was partially provided by the Human Frontier Science Program grant RG0109/1999-B, AFOSR grant 93NL366 and NIH grant EY08266.

References

[1] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (Second Ed.). New York, NY: Springer-Verlag.

[2] Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science*, *4*, 99–108.

[3] Freeman, W. T. (1996). The generic viewpoint assumption in a bayesian framework. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 365–389). Cambridge, UK: Cambridge University Press.

[4] von Helmholtz, H. (1865–67). *Handbuch der physiologischen Optic*. Hamburg, Germany: L. Voss.

[5] Hershenson, M. (1999). *Visual Space Perception: A Primer*. Cambridge, MA: MIT Press.

[6] Kersten, D. (1999). High-level vision as statistical inference. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.

[7] Knill, D. C., Kersten, D. & Yuille, A. L. (1996). Introduction: A bayesian formulation of visual perception. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 1–21). Cambridge, UK: Cambridge University Press.

[8] Knill, D. C. & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge, UK: Cambridge University Press.

[9] Maloney, L. T. (2001). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the Physical World*. Chichester, UK: Wiley.

[10] Mamassian, P. & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*, B1-B9.

[11] Mamassian, P. & Landy, M. S. (1998). Observer biases in the 3d interpretation of line drawings. *Vision Research*, *38*, 2817–2832.

[12] Mamassian, P. & Landy, M. S. (in press). Interaction of visual prior constraints. *Vision Research*.

[13] Myers, J. L. (1976). Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of Learning and Cognitive Processes: Approaches to Human Learning and Motivation*, Volume 3 (pp. 171–205). Hillsdale, NJ: Erlbaum.

[14] Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society*, *2*, 37–42.

[15] Sun, J. & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, *1*, 183–184.

[16] Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

[17] Szeliski, R. (1989). *Bayesian Modeling of Uncertainty in Low-Level Vision*. Boston, MA: Kluwer.

[18] Yuille, A. & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 123–161). Cambridge, UK: Cambridge

University Press.

[19] Zhu, S. C. & Mumford, D. (1997). Prior learning and gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 1236–1250.