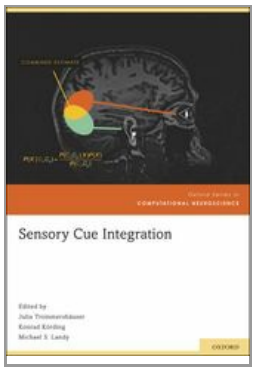

Sensory Cue Integration
Julia Trommershäuser, Konrad Kording, and Michael S. Landy

Print publication date: 2011
Print ISBN-13: 9780195387247
Published to Oxford Scholarship Online: September 2012
DOI: 10.1093/acprof:oso/9780195387247.001.0001

The Role of Generative Knowledge in Object Perception

Peter W. Battaglia
Daniel Kersten
Paul Schrater

DOI:10.1093/acprof:oso/9780195387247.003.0003

[–] Abstract and Keywords

Combining multiple sensory cues is an effective strategy for improving perceptual judgments in principle, but in practice it demands sophisticated processing to extract useful information. Sensory cues are signals from the environment available through sensory modalities; perceptions are internal estimates about the world's state derived from sensory cues and prior assumptions about the world. The influence that world properties have on sensory cues is inherently complicated, so recovering information about the world from sensations is a difficult problem. This chapter discusses “generative knowledge” as a unifying framework regarding how biological brains overcome these difficulties to interpret sensory cues.

Keywords: sensory cues, perceptual judgments, perception, generative knowledge

INTRODUCTION

Combining multiple sensory cues is an effective strategy for improving perceptual judgments in principle, but in practice it demands sophisticated processing to extract useful information. Sensory cues are signals from the environment available through sensory modalities; perceptions are internal estimates about the world's state derived from sensory cues and prior assumptions about the world. The influence that world properties have on sensory cues is inherently complicated, so recovering information about the world from sensations is a difficult problem. Imagine trying to analyze measurements from a scientific experiment without knowing the experimental protocol: It is impossible to draw conclusions without knowing the methods by which the raw data were produced. Likewise, converting raw neural sensory signals into perceptual judgments, like the distance to a nearby object or the material it is made of, requires the application of knowledge that is both structured and flexible. In this chapter, we discuss "generative knowledge" as a unifying framework regarding how biological brains overcome these difficulties to interpret sensory cues.

Challenges for Perception

Perceptual processes translate raw sensory data into high-level interpretations. In doing so, perception solves several major computational challenges:

1) The mapping from sensory data to interpretations can be complex and ill posed. For example, extracting three-dimensional (3D) geometry from two-dimensional (2D) images is fundamentally ambiguous because a multitude of 3D structures can project to any single 2D pattern (Marroquin, Mitter, & Poggio, 1987). More generally, the relationship between each sensory cue and the environmental properties that caused it may be intricate and convoluted (the acoustic vibrations arriving at the eardrum or the pattern of light intensities on the retina have no simple, unambiguous relationship with the position of an object's sound source or its geometric shape).

2) Many sensory cues do not relate directly to the environmental properties of interest, but rather contain "auxiliary" information related to the quality and meaning of other cues. For instance, when judging the distance to a face, knowing its physical size is **(p.47)** irrelevant in isolation, but it can be used to disambiguate the visual image size (Yonas, Pettersen, & Granrud, 1982). Employing auxiliary information is important, but it requires understanding of how the multiple cues are related to each other.

3) Sensory cues vary in quality:

i) Relative to each other. For instance, vision provides higher spatial resolution than audition, whereas audition provides higher temporal resolution than vision.

ii) Depending on external factors. For instance, in fog visual cues may provide worse spatial information than auditory cues.

iii) Depending on internal factors. For instance, cataracts and uncorrected myopia diminish visual acuity.

iv) As a function of the world state. For instance, binocular stereo cues to

slant decrease in reliability as surface slant increases, whereas texture compression cues to slant increase in reliability as the slant increases (Knill, 1998a, 1998b).

Because of variability in the quality of cues, the brain must know when, and how much, to trust cues' information and when they are too unreliable to be informative (see Chapter 1; Ernst & Banks, 2002; Jacobs, 1999).
4) The arrangements of objects'spatial and material properties in the world follow highly predictable patterns. Though it's possible for objects to appear in an infinite variety of arrangements, we only ever encounter a very small subset of the possible configurations. For instance, water faucets almost exclusively appear in bathrooms and kitchens near waist level. In the absence of particularly strong sensory evidence to the contrary, such knowledge can be used to immediately exclude perceptual scene interpretations that involve a water faucet on the ceiling. Statistical distributions about object properties and context can be learned; these distributions are termed *priors*. Priors offer tremendous benefits for perception by helping overcome ambiguity and impoverished sensations, but because of the vastness and complexity of scene knowledge, knowing how to organize and use it is neither obvious nor trivial (Strat & Fischler, 1991).

Overcoming Perception's Challenges Using Generative Knowledge

The challenges presented earlier are a consequence of the complex relationships among the set of immediate world properties and the sensory cues they generate. The brain draws percepts from sensations by taking advantage of knowledge about these complex relationships. To study how observers use their knowledge for perception, it is useful to precisely identify the potential sources of this knowledge.

We use the term *sensory generative process* to characterize how sensations are caused by the world. This includes the physical factors that lead to the stimulation of the sensory organs, and the relationships among world properties that make some situations more common than others. Some examples include the optical projection process by which light stimulates the retina, and the typical arrangements of objects in an office that favor large objects being placed on the floor and smaller objects being placed on desks and shelves. In general, the generative process refers to those events that happen before the sensations arrive at the brain and which constrain sensory input in a physically predictable manner.

The term *sensory generative knowledge* refers to built-in assumptions held by the brain about the sensory generative process—it links sensory cues back to the world properties that caused them. An observer who interprets sensory cues in the context of their generative process can make more accurate judgments about the world by combining sensory cues and prior information to constrain possible scene interpretations, which leads to more accurate and robust perceptions.

As an example, consider the relationship between retinal image size and object size. An object's image size on the retina is influenced by two factors, the object's physical size

and distance. Image size alone does not allow an observer to unambiguously determine the size or distance: The object could be small and near, or large and far; either situation **(p.48)** may produce the same image size (Fig. 3.1A). Now, consider the sensory generative process. Because of perspective projection, two factors play a dominant role in determining monocular image size (I) (measured in visual angle): the object's physical size (S) and distance (D). The generative relationship between S, D, and I can be summarized by the function

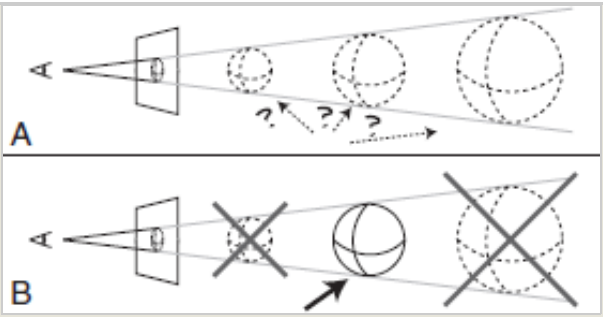

Figure 3.1 (A) The size of an object's image does not uniquely specify its physical size and distance, only a set of size and distance combinations consistent with the image. (B) If one knows the distance, the size can be uniquely determined; likewise, if one knows the size, the distance can be uniquely determined.

$$\frac{S}{D} = I.$$

(3.1)

For an observer that has this generative knowledge and can measure I, several facts about S and D are immediately apparent. First, it is possible to solve for the precise value of S only if D is known, and it is possible to solve for D precisely only if S is known: It is not possible to solve one equation for two unknowns. However, if an auxiliary cue to size or distance is available, estimating the other becomes possible (Fig. 3.1B). Second, if information about either S or D is available, but uncertain, this provides uncertain information about the other. For instance, if you are told that S is between 1 and 2 meters in diameter, and I is 1/10 radians, you can infer that D is between 10 and 20 meters.

This simple illustration shows how generative knowledge can help overcome the ill-posed, ambiguous nature of perception through auxiliary cues (challenges 1 and 2 presented earlier). When judging an object's distance, cues to its physical size, although independent of distance, can nonetheless be used to disambiguate the causes of the extent of the image on the retina.

Sensory cues vary in quality (challenge 3), and an observer who knows the relative reliability of available sensory cues can differentially incorporate cues of varying quality to form more accurate perceptions. Consider an experiment in which an object is viewed

binocularly and the observer also reaches out and touches it. Thus, at least two cues are available to the object distance: vergence angle and felt arm position while touching the object (haptic cue). If one cue were less reliable than the other, it should be allowed less influence on the perceptual judgment. For instance, if vergence angle is a more reliable cue to the distance, then in an experiment in which the two cues are set in conflict (meaning they indicate different distances), the observer's perceptual distance judgment would more closely reflect the vergence-indicated distance than the haptic-indicated distance.

Exploiting knowledge of how world properties relate to other world properties (challenge 4) can be useful for constraining possible perceptual interpretations. In the size/distance example, if the observer recognizes that the object whose size is being judged is a face, this provides a strong restriction on possible sizes because the variance among face sizes is very small. Alternatively, if the observer is trying to judge the distance to the object, the prior knowledge that face sizes fall in a tight range can be used to rule out size/distance combinations that are inconsistent with the size prior. More generally, almost every perceptual behavior is heavily influenced by contextual information (Biederman, 1772; Oliva & Torralba, 2007). For instance, a large, horizontally extended object on a street is likely a car, whereas a vertically extended object indoors is likely a person.

Generating Perceptual Samples

Although some kinds of generative knowledge can be embodied in a purely feedforward **(p.49)** inference process, mechanisms that use generative knowledge to hypothesize new instances that have never been experienced can provide considerably more flexible and robust visual inferences (Yuille & Kersten, 2006). Such generative models of the world provide a functional link between world properties and sensory cues that can take a hypothesized world property and compute its probable sensory consequences. Such generative knowledge is believed to form a critical part of the motor control system, called a *forward model* (Kawato, 1999), that predicts the afferent sensations that will result from motor commands. Much less is known about the existence of sophisticated generative models for perception; however, neural predictive-coding models posit that the brain has higher level perceptual processing sites that generate predictions of sensory input at lower levels (Mumford, 1992; Rao & Ballard, 1999). In addition, humans' dreams and visual imagery abilities suggest generating perceptual samples is possible. For example, imagine you are looking at a kitchen. Immediately items like refrigerators, stoves, sinks, and tables come to mind, and you can provide likely colors, sizes, and positions of such objects. The particular kitchen you conjure may be a kitchen you have seen in the past, but it is also possible to imagine a new kitchen you have never before seen. The ability to visualize and imaginatively construct unobserved scenes may reflect the operation of complex generative models of the visual world. Lastly, predicting causal chains of events, like a tennis ball's future position after a series of bounces, can be aided by an approximate generative model of elastic collisions and momentum.

In the next (second) section, we introduce the theoretical issues concerning generative

knowledge in the context of Bayesian inference and the joint roles of sensory cues and prior knowledge for perceptual inference. In the third section, we present empirical results that support the brain's use of sensory generative knowledge. In the fourth section, we discuss inference of world properties when nuisance properties confound available cues.

THE BAYESIAN OBSERVER MODEL

Background

Bayesian inference is a model for perception that has achieved broad support for several reasons (Knill & Pouget, 2004; Körding & Wolpert, 2006). Methodologically, it is a principled, rigorous language for probabilistic models suitable for characterizing perceptual inference based on sensations and prior knowledge. As described earlier, perception solves the frequent problem of ambiguity due to the noninvertibility of sensations; formally, inference is a process of inversion under uncertainty in which a set of possibilities are obtained, rather than a unique solution. Bayesian models naturally describe the combination of prior knowledge and available sensory cues, as well as how observers learn from experience to make better-informed judgments in the future, both of which are common phenomena in biological perception. Through Bayesian models, numerous studies have reported optimal and near-optimal performance across a gamut of perception tasks (for a review, see Kersten, Mamassian, & Yuille, 2004). Of equal importance, failures of optimality provide an opportunity for identifying limitations in neural processing and deviations between human and model assumptions (see Chapter 8).

Perception as Bayesian Inference

As presented in Chapter 1, Bayes' rule specifies how to optimally combine measurements and prior information to gain information about unobserved quantities, a computation termed *Bayesian inference*. In biological perception, the brain directly measures sensory cues but does not directly measure external world properties. By treating both cues and world properties as *random variables*, and quantifying their respective conditional and marginal probability distributions, Bayesian inference provides a probability distribution over possible world states that can be used to make optimal scene estimates. Bayesian models provide a **(p.50)** powerful normative framework for describing and evaluating theories of perceptual behavior.

Let the relevant state of the world, or those properties the observer is interested in, be represented by R, and direct sensory measurements by D. Bayes' rule specifies:

$$P(R|D) = \frac{P(D|R)\,P(R)}{P(D)},$$

(3.2)

where $P(R|D)$ is the *conditional likelihood* of D given R, $P(R)$isthe *prior probability* of R, $P(D)$ is the *marginal likelihood* of D, and $P(R|D)$ is the *posterior probability* of R given D.

The term $P(D|R)$ represents the generative relationship between world properties R and sensory cues D.

We can use these variables to represent elements from the size- and distance-perception example in the previous section. Consider an observer who is trying to judge the distance to a ball; we represent distance as the relevant variable, R. Assume the observer reaches out and touches the ball, so that the felt arm position provides a direct distance cue, D. We represent the conditional relationship of the sensory cue given the relevant world property as $P(R|D)$ (Fig. 3.2, dashed curve). We represent prior knowledge about different ball distances as $P(R)$ (Fig. 3.2, dotted curve). For this example the prior specifies the ball's probable distance before considering specific sensory cues; in this case, it may reflect knowledge that the ball is between 1 and 2 meters from the observer. The term $P(D)$ characterizes the probability of receiving cue D. For any particular sensory cue, $P(D)$ is constant and because the right-hand side is a probability distribution (that integrates to one), $P(D)$ is fully determined by $P(D|R)\,P(R)$. Bayesian inference proceeds by merging the cue likelihood and distance prior to form a posterior probability distribution over the possible ball distances.

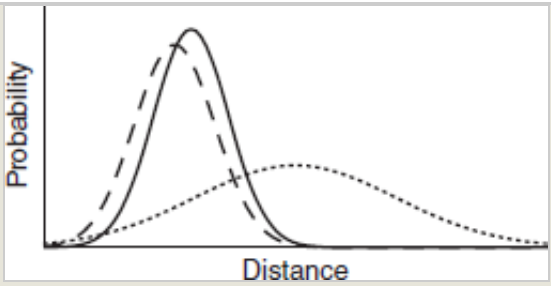

Figure 3.2 The dashed curve represents the conditional likelihood of a sensory cue given different distances. The dotted curve represents the prior probability distribution over different distances. The solid curve represents the posterior probability distribution over different distances given the cue; note that its peak lies between the likelihood and prior peaks—closer to the peak specified by the more reliable, likelihood function.

Bayes' Nets

Pearl (1988) introduced Bayes' nets (see examples in Fig. 3.3) as directed, acyclic graphical models that express the conditional probability relationships among multiple random variables and their prior probability distributions. Bayes' nets are a useful tool for describing sensory generative processes and the requisite inference rules because they allow graphical expressions of properties that otherwise must be represented by dense symbolic notation. And perhaps more important, they allow similarities among seemingly unrelated sensation-perception behaviors to be recognized by the modeler. The circles, called *nodes* (Fig. 3.3), represent random variables. Nodes that have no parents, termed *roots,* have prior probability distributions over their possible values. The arrows connecting the nodes, called *edges*, represent conditional probability distributions among random variables. In sensory generation, conditional dependencies

(edges) typically represent causal relationships. For instance, the node labeled *R* in Figure 3.3A can represent the ball's distance and *D* can represent the haptically sensed arm position cue from the size-distance example in the previous subsection. The arrow connecting them represents the conditional likelihood function of sensory cue given distance. In general, the direction of an edge is arbitrary (the edge in Fig. 3.3A represents $P(D|R)$, but if its direction were reversed would represent $P(R|D)$). Modelers choose directions that best suit the problem and system being modeled. For perceptual inference, the generative process relates to the forward direction from world to sensations; the inference process is the reverse **(p.51)** **(p.52)** direction, from sensory cues to world properties. As a terminology note, we use the term *direct* to refer to cues that are connected to relevant states by an edge. In contrast, the term *auxiliary* is used to refer to cues that do not share an edge with any relevant states but are connected by a sequence of edges between intermediate nodes.

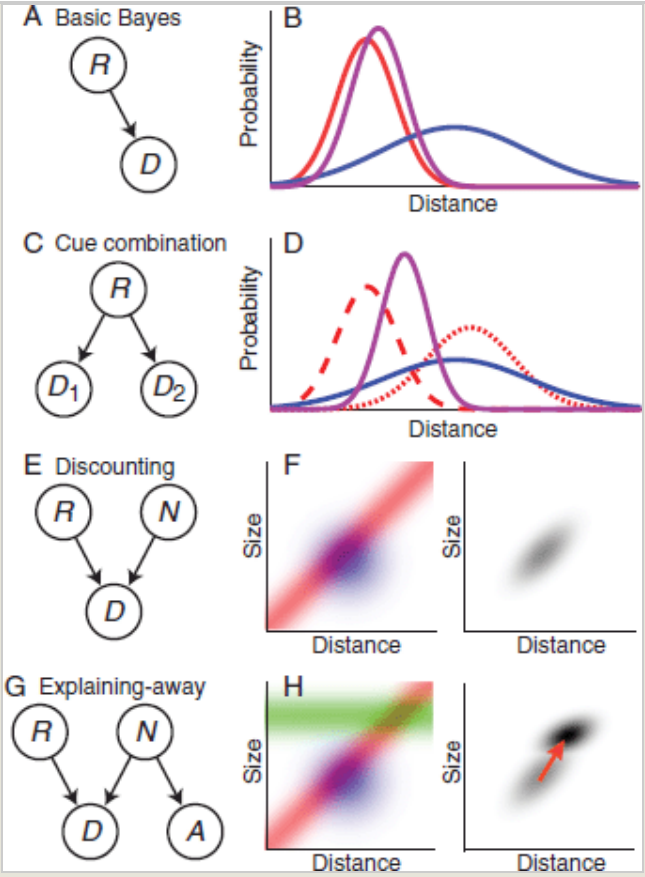

Figure 3.3 (A) Basic Bayes: A relevant world property, *R*, causes a direct cue, *D*. (B) The optimal inference strategy is to combine cue likelihood (red curve) with the prior assumption about the relevant property (blue curve) to compute a posterior probability distribution (purple curve). Note that the prior probability distribution is over the root node, *R*. (C) Cue combination: A relevant world property, *R*, causes two likelihood (red dashed/dotted curves) with the prior assumption about the relevant property (flat, blue curve) to compute a posterior probability distribution (purple curve). Bayes' rule prescribes calculating their product. (E) Discounting: Two world properties cause a direct cue, *D*. The observer needs to perceive the relevant world property, *R*,

and not the nuisance property, N, but R is ambiguous given only the sensory cue. (F) The optimal inference strategy is to combine the cue likelihood (red distribution, left) with the prior assumptions about the nuisance and relevant properties (blue distribution, left) to compute a posterior probability distribution (gray distribution, right). Bayes' rule prescribes calculating their product. The prior assumptions disambiguate the sensory measurement. (G) Explaining away: Two world properties, relevant, R, and nuisance, N, cause a direct cue, D, and auxiliary cue, A. The observer needs to perceive R, but not N, but R is ambiguous given only the cue. Though A is not directly related to R, it provides information about N that can disambiguate R. (H) The optimal inference strategy is to combine the direct (red distribution, left) and auxiliary cue likelihoods (green distribution, left) with the prior assumptions about the nuisance and relevant properties (blue distribution, left) to compute a posterior probability distribution (black distribution, right). The prior assumptions and auxiliary cues disambiguate the sensory measurement. The arrow shows how the discounting posterior distribution (panel F) shifts (as well as tightening to become more peaked) as a result of the auxiliary cue.

Generative Knowledge in Bayesian Inference

Figure 3.3 illustrates four elementary sensory generative process models (Kersten et al., 2004). These are not unique or exclusive; there may be more than one way to characterize a sensory circumstance. Instead they can be thought of as choices made by the modeler to characterize particular sensation-perception events in an accurate yet succinct manner. As an extreme example, ambient air temperature affects the index of refraction between air and eye, and thus influences the visual generative process, but the impact is negligible so the modeler can choose to ignore it.

Figures 3.3A and 3.3B represent situations in which a single world property causes a single cue and follows a relatively straightforward generative process. Some examples include a moving object producing a moving image on an observer's retina, the distance to an object being measured by binocular vergence, and the position of a sound source being sensed through interaural auditory differences. Inference in this situation, termed *basic Bayes*, is performed by inverting the nondeterministic functional relationship between the cue and world property, and combining this information with prior knowledge about the world property like the example in the previous subsection.

Figures 3.3C and 3.3D represent situations in which a single world property causes multiple cues. Some examples include a surface producing binocular stereo and texture compression cues to its slant, the distance to an object being measured by binocular vergence and felt arm position, and the position of a sound source being sensed through interaural auditory differences and visual cues. Inference here, termed *cue combination*, is similar to the basic Bayes case. It is performed by inverting the direction of influence between each cue and world property, and combining these inverted relationships with prior knowledge about the world property.

Figures 3.3E and 3.3F represent situations in which multiple world properties influence

one cue. Some examples include illuminant intensity and surface reflectance causing a sensed luminance cue, or an object's size and distance each influencing its monocular image size. When an observer infers one (relevant) world property among other, nuisance properties, termed *discounting,* the cue only can constrain the possible relevant property values to a set of relevant-/nuisance-value combinations. Prior knowledge about the nuisance property must be used to rule out unlikely relevant/nuisance combinations.

Figures 3.3G and 3.3H represent situations in which multiple world properties influence multiple cues; the cues can be divided into those that are *directly* influenced by the relevant world property, and auxiliary ones that are only *indirectly* related to the relevant world variable. Some examples include a surface's shape and reflectance each influencing a sensed luminance cue and the shape also influencing a visual geometry cue, an object's size and distance each influencing its sensed image size and the distance also influencing a binocular vergence cue, and two spatially separated sound sources each causing interaural auditory difference cues and one source also causing a visual cue to its position. When an observer infers one (relevant) world property among other, nuisance properties, termed *explaining away,* the direct, confounded cue only can constrain the possible relevant property values to a set of possible relevant/nuisance value combinations (as in discounting). However, auxiliary cues (those not directly related to the relevant world property) and prior knowledge about the nuisance property can be used to rule out unlikely relevant/nuisance combinations.

The conditional-likelihood and prior-probability terms implicitly dictate how strongly the sensory cues and prior knowledge should influence the final perceptual inference. When sensory information propagates backward through the generative structure **(p.53)** in inference, the uncertainty in the conditional distribution determines the relative impact of the information: For conditional dependencies with low uncertainty the information is very influential; for high uncertainty the information plays a lesser role. The same is true for the uncertainty in prior probability distributions.

Discounting and Explaining Away

Discounting and explaining-away inference processes critically depend on generative knowledge. It is easier to conceive noninferential, associative learning systems that conduct cue-combination-like inference, even using relative cue reliability, but it is difficult to contrive reasoning patterns like “explaining away” without generative knowledge and Bayesian inference. Thus, discounting and explaining-away phenomena form stronger tests of humans' use of generative knowledge for perceptual inference than cue combination. However, new analysis tools must be developed for testing discounting and explaining-away phenomena that entail more complex ideal-observer models. The fourth section presents a novel framework for analyzing more complex ideal-observer models.

The following section reviews qualitative reports of perceptual discounting and explaining away.

EXPERIMENTAL EVIDENCE FOR THE USE OF GENERATIVE KNOWLEDGE: DISCOUNTING AND EXPLAINING AWAY

Observers frequently receive ambiguous sensory input, which makes interpreting the scene challenging because more than one possible interpretation could be correct. Perceptual discounting and explaining away are behaviors that overcome this problem using generative knowledge.

Discounting

Studies of perceptual discounting have found evidence that shape-from-shading perception is influenced by prior assumptions about illuminant direction in accordance with the generative relationship between shape, illumination, and sensed luminance. Mamassian and Goutcher (2001) measured human observers' estimates of an object's shape from shading cues, which require an assumption about what direction the light arrives from, to be useful. This prior knowledge helps to disambiguate the otherwise ambiguous shading cue. Mamassian and Landy (2001) investigated how multiple priors' weights are decided by the brain, specifically lighting direction and surface slant priors, and concluded that the weights reflect their relative reliabilities. Adams, Graf, and Ernst (2004) modified observers' light direction priors by providing haptic feedback that suggested a different lighting direction than the default overhead assumption.

Explaining Away

Some of the most striking examples of perceptual explaining away can be demonstrated with ambiguous, especially bistable, stimuli. Bistable stimuli are those that have more than one perceptual interpretation, and when viewing the stimuli the perceptual experience spontaneously “flips” between interpretations with a period of roughly 5–45 seconds, though sometimes much longer. Examples include the “Necker cube” and kinetic-depth-effect rotating cylinders (see Chapter 9 for a picture). Studies have shown (Blake, Sobel, & James, 2004; James & Blake, 2004) that by providing an auxiliary sensory cue, like binocular stereo or haptic input, the bistability can be reduced or removed altogether. This auxiliary cue serves to explain away particular stable interpretations that are inconsistent with the cue.

Knill and Kersten (1991) reported a clear instance of perceptual explaining away in which an object's surface shape affects observers' judgments of albedo, consistent with a generative model that explains away the effect of the shape on the luminance. Figure 3.4 illustrates Knill and Kersten's (1991) stimuli in which generative knowledge allows auxiliary shape cues to disambiguate an otherwise ambiguous luminance cue to the albedo of a surface (adapted from Knill & Kersten, 1991). In the upper row, the grayscale image shows two objects **(p.54)** with different shapes. The observer's task is to decide what the albedo is in a horizontal cross-section across each object (dashed white boxes). Under the image, the rows represent the actual luminance profile of the pixels across each object (labeled “L”), and the perceived albedo profile across each object for a typical observer (labeled “A”). The perceived albedo profiles are due to the different perceived shapes of the objects, indicated by their respective edge cues. In the left object, because the shape looks flat on the front, the luminance difference across the

object's center is attributed to variation in albedo. In the right object, the left side of each cylinder is perceived to face the light source more directly; changing albedo is not required to explain the luminance differences across the center of the object.

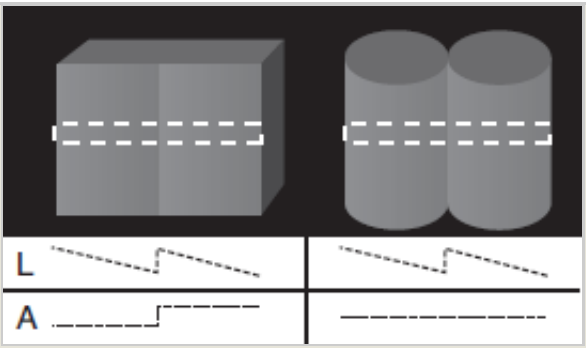

Figure 3.4 The image depicts two objects with different shapes, yet the luminance profiles within the dotted boxes are identical for both objects, as plotted below the objects in the row labeled "L". The perceived albedo, plotted in the row labeled "A", is different for the different objects.

Figure 3.3G shows the graphical model that characterizes this perceptual-inference problem. The relevant property (R) represents the object's albedo/reflectance, the nuisance property (N) represents the object's surface shape, the direct cue (D) is the luminance, and the auxiliary cue (A) is the object's boundary contour that provides direct information about surface shape. Luminance is a function of both surface shape and albedo (and light source intensity and direction, too, but here we assume they are constant). The observer's perceptual task is to estimate the albedo, but the effect that shape has on luminance must be explained away in order to disambiguate the albedo. Because there is auxiliary boundary contour information that provides an independent estimate of the shape, the shape's impact on luminance can be explained away and the albedo unambiguously estimated. The observer requires knowledge of how shape and albedo generate the image data (boundary contour and luminance) to use it for perceptual estimation.

Size- and Distance-Perception Experiments

The influence of auxiliary distance cues on human size perception has received much attention for more than a half century. However, few studies measuring the influence of size cues on distance perception have been reported, and no experiments have investigated how cues to an object's changing distance influences perception of whether, and to what degree, its size changes.

Holway and Boring (1941) found size constancy (in which perceived size is proportional to object size) was best facilitated by providing many, strong distance cues, though others (Epstein, Park, & Casey, 1661; Gogel, Wist, & Harker, 1663; Ono, 1966) concluded that size constancy was subject to a variety of failures. Epstein et al. (1961) and other authors (Brenner & van Damme, 1999; Gruber & Dinnerstein, 1665; Heinemann & Nachmias, 1665; Ono, Muter, & Mitson, 1974) acknowledge that distance

judgments are not always veridical (apparent distances do not always match physical distances), which accounts for some size mis-perceptions, and specific experimental design choices and task demands often contribute to the nature of the experiment's recorded failure of size constancy (Blessing, Landauer, & Coltheart, 1667; Kaufman & Rock, 1662; Mon-Williams & Tresilian, 1999).

Several studies investigated humans' use of size information for making distance judgments. Granrud, Haake, and Yonas (1985) showed that 7-month-old infants who were allowed to learn the size of different objects by playing with them used the size to judge the distance in postplay test phases. In contrast, 5-month-old infants did not exhibit the use of size information for distance judgments, suggesting **(p.55)** the development of knowledge about size and distance occurs as early as 5–6 months old. Yonas, Granrud, and Pettersen (1985) showed that when presented with two objects of different retinal visual angles, infants older than 5 months perceived the larger as nearer, but not 5-month-olds. Yonas, Pettersen, and Granrud (1982) showed that 7-month-old infants' and adults' distance judgments are influenced by familiar-size information associated with faces, but 5-month-olds are insensitive to familiar size. These results suggest that size information can influence distance judgments.

We now describe results from two experiments, one investigates how sensory measurements of depth changes influence judgments of physical object size changes, and a second in which measurements of size influence a depth-dependent action.

The role of auxiliary distance information in size perception was addressed by a recent study. Battaglia et al. (2010) conducted an experiment in which they presented participants with balls that moved in depth and simultaneously either inflated or deflated, and they asked participants to decide “inflation” or “deflation” for each stimulus. The experimenters provided binocular and haptic cues to the ball's distance change to test the effect these auxiliary sensory cues had on size-change perception. They reasoned that because objects do not usually change in size, if participants make use of the auxiliary cues they must have general knowledge of the relationship between size and distance, and not simply exploit a learned association between the auxiliary sensory cues and size-change perception. The results were that in the absence of auxiliary cues, participants relied on prior assumptions that the object was stationary to judge the size change proportional to the image size change, and they made perceptual mistakes in cases when the distance change had a large, opposite effect on image size than the physical size change. But both the binocular and haptic cues were effective in nulling that bias by providing disambiguating distance-change information. Interestingly, binocular cues were more effective than haptic cues, which may reflect observers' weaker “trust” of haptic cues due to possible dissociation from the visual object. They concluded that humans must have knowledge of the relationship between size, distance, image size, and the auxiliary distance cues to make these perceptual judgments.

Figure 3.5 depicts one participant's data in Battaglia e al. (2010)'s experiment; H refers to “haptic” auxiliary cues, B means “binocular” auxiliary cues, a plus sign means the cue was present, and a minus sign means the cue was absent. Notice that in the case with no

haptic and no binocular auxiliary cues (labeled H–/B–), those stimuli perceived as "inflating" (gray region) were predicted by whether the image size was growing or shrinking (black, diagonal, dashed line). When haptic and/or binocular cues were available (labeled H+/B–, H–/B+, H+/B+), participants' perception of inflating balls changed to reflect the true physical size change more accurately.

Battaglia, Schrater, and Kersten (2005) conducted an experiment in which participants were asked to intercept a moving ball that varied in size across trials, by positioning their hand at a distance of their choice. The participant's hand was constrained so that it could only move along the line of sight, and the ball moved from left to right, crossing the hand's constraint line at variable distance. The hand's distance placement was considered to be a measure of the participant's percept of the ball's distance. In some trials the participants were allowed preinterception haptic interaction with the ball, which provided an auxiliary cue to the ball's size. By comparing participants' distance judgments in the "haptic auxiliary cue" condition with those in the "no haptic auxiliary cue" condition for trials with identical ball image sizes, experimenters were able to measure participants' abilities to explain away the confounding influence of the ball's physical size on the image size. The distance judgments of an "explaining-away observer" should be less dependent on the physical size than the distance judgments of an observer with no auxiliary information (for one participant, Fig. 3.6). Figure 3.7 summarizes all participants' results, which support the hypothesis that participants explain away the influence of physical size when making distance **(p.56)** judgments. In particular, Figure 3.7A depicts the correlation between participants' distance judgments and the balls' physical sizes and shows that participants' distance judgments were less dependent on the ball's physical size when auxiliary size cues were available to explain away the size confound. Figure 3.7B shows that interception performance improved as a result of this explaining-away reasoning.

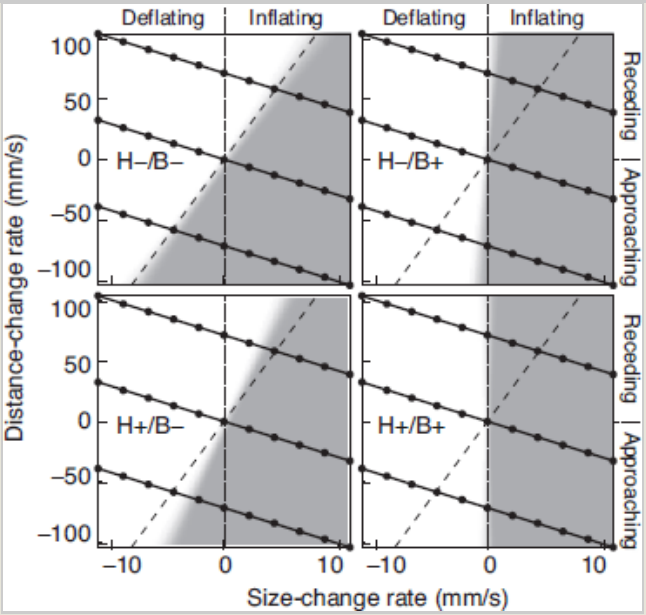

Figure 3.5 One participant's judgments of balls as inflating (gray) or deflating (white) balls. Each box represents a unique combination of haptic and binocular distance cues (indicated by "H* / B*" on the left side of each box). The black dots represent different size- and distance-change stimulus values. The black diagonal dashed line represents those stimuli whose image size did not change; left of the line indicates shrinking image sizes, and right of the line indicates growing image sizes. The black vertical dotted line indicates the true boundary between inflating and deflating balls. We interpolated between the 50% points of psychometric functions across the solid diagonal lines to estimate the gray/white, inflation/deflation boundaries. When haptic and binocular distance cues are available, the participant's judgments of which stimuli were inflating became more accurate because the confounding influence of distance on image size was explained away by the auxiliary distance cue.

INFERENCE IN THE PRESENCE OF NUISANCE WORLD PROPERTIES

Until now our discussion has dealt with normative models of perceptual inference, and we have presented *qualitative* evidence that observers use generative knowledge to make perceptual judgments. To model perceptual inference *quantitatively*, a Bayesian observer model is required. We now describe how to use a behavioral experiment to test, and estimate parameters of, such a model.

It is important to realize that for more complex perceptual-inference situations, like discounting and explaining away (Fig. 3.3), the observer requires generative knowledge to interpret input sensations and prior knowledge. An important question is: What generative knowledge does the human observer possess? On the one hand, it seems unlikely that observers know the exact nature and quality of each sensory cue's relationship with the world. On the other hand, human observers' excellent perceptual performance across a wide range of tasks suggests that they use very sophisticated strategies that **(p.57)** **(p.58)** may include detailed internal knowledge of generative processes. The remainder of this section describes a formal framework to analyze this

question.

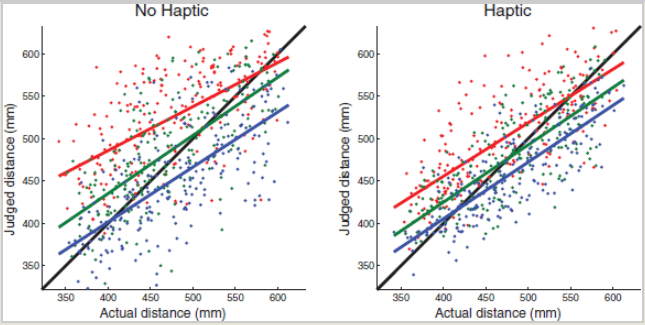

Figure 3.6 These scatter plots show a typical participant's interception behavior. Each dot represents a single trial's data point; red were smaller balls, green were medium balls, and blue were large balls. The x-axis shows the actual distance of the ball. The y-axis shows the participant's judged distance. The black diagonal line shows the line that indicates perfect judgments. The colored lines are regression fits to the small, medium, and large balls, with color corresponding to the dot colors. The left figure shows the no-auxiliary-haptic-size-cue condition; the right figure shows the condition in which the auxiliary haptic size cue was provided.

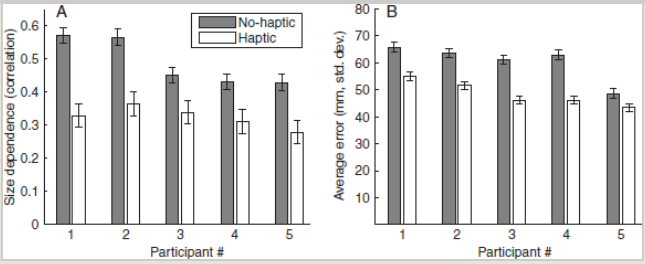

Figure 3.7 (A) Each pair of bars represents the correlation between distance judgments and ball size for one participant; the gray bars depict the "no-auxiliary-size-cue" condition, and the white bars depict the "haptic-auxiliary-size-cue" condition. All participants' distance judgments are less correlated with the physical size cue when auxiliary information is available to disambiguate the image size cue. (B) Each pair of bars represents participants' average error (standard deviation of distance judgments). Auxiliary cues improved all participants' performance.

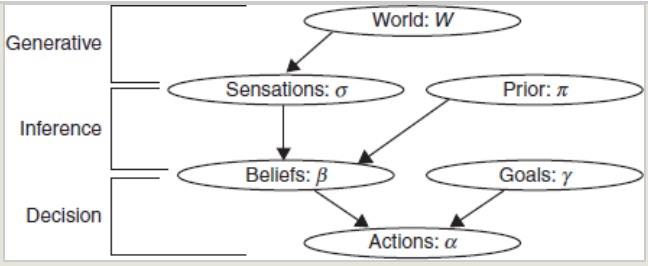

Figure 3.8 Full psychophysical-observer model. The sensation process is characterized by the sensory generative process in which world properties (W) produce sensory input (σ) in the observer. Perception is characterized by the inference process in which the observer combines sensory cues (σ) with prior knowledge (π) to compute beliefs (β) about the state of the external environment (W). Actions are characterized by the decision-making process in which the observer combines beliefs (β) with goals (γ) to select actions (α) that are predicted to result in highest reward. This model can be formally quantified using the Bayesian decision-theoretic formalism, provided in Eq. 1.5 in Chapter 1.

Generative Processes versus Knowledge

The distinction between the sensory generative process and an observer's generative knowledge must be recognized when modeling an observer's performance in a psychophysical task. The generative process entails how the world produces sensory cues through physical processes and the statistical regularities of natural scenes that are external to the observer. Generative knowledge is the observer's model, or understanding, of the generative process that is either built in or acquired through experience. For *subjectively* optimal Bayesian inference, the observer uses its generative knowledge in accordance with Bayes' rule for inference. However, for the stronger case of *objectively* optimal Bayesian inference, the generative knowledge also accurately reflects the true generative process.

Figure 3.8 represents a model observer who receives sensory cues from the world (labeled "generative"), integrates prior assumptions with those cues to form perceptual inferences (labeled "inference"), and selects actions based on those inferences and its internal goals (labeled "decision"). Though it does not capture phenomena like attention, sensorimotor feedback, or learning, this framework is very useful for quantifying how a psychophysical observer produces responses based on input stimuli.

The state of the world is represented by the top node and is labeled W; the actor's sensory cues are labeled σ; the actor's prior knowledge is labeled π; the actor's inferred state of the world, or *beliefs*[1] about the world, are labeled β; the actor's goals are labeled γ; and the actor's actions are labeled α. The arrow from $W \rightarrow \sigma$ represents the sensory generative process; the arrows from $\sigma \rightarrow \beta \text{and} \pi \rightarrow \beta$ represent the perceptual inference process, which is guided by generative knowledge; and the arrows from $\beta \rightarrow \sigma \text{and} \gamma \rightarrow \alpha$ represent the decision process by which the actor chooses actions in response to his or her beliefs about the world and internal goals. The actor has access to sensory cues, σ; prior knowledge, π; beliefs about the world, β; and goals, γ.

Although the actor controls his or her actions, the actual outcome of the actions, α, varies with respect to the intended behavior.

(p.59) The experimenter has access to the world state, W, and action measurements, α.

This modeling framework allows an experimenter to manipulate W in order to study the generative, inference, and decision processes within observers. An ideal-observer model (see Chapter 1) can be used to parameterize the experimentalist's assumptions and hypotheses about the generative, inference, and decision processes for a psychophysical observer, and to compute predicted actions α, given W. In this way behavioral measurements can be directly compared to model predictions and used to estimate the model's parameter values. When observers' behaviors outperform suboptimal models, these models can be immediately dismissed. Observers' deviations from optimality (see Chapter 8) suggest their use of heuristics, which may be more easily pinpointed by considering the ideal-observer model. Additionally, because of the relationship between human generative knowledge and the true generative process, the estimated *generative knowledge* model parameters can be compared to the estimated *generative process* parameters to assess the quality of the observer's knowledge. Generally ideal-observer predictions allow experimentalists to classify observers into three groups: (1) those who have accurate generative knowledge and make optimal perceptual inferences (objectively optimal observer), (2) those who are suboptimal because they apply inaccurate generative knowledge in a Bayes-consistent manner (subjectively optimal observer), and (3) those who are suboptimal because they do not draw perceptual conclusions in accordance with Bayesian inference rules at all.

This modeling framework also allows the experimenter to ensure a proposed experiment has sufficient power to adequately test a hypothesis. Often there are many assumptions and unknown parameters in a model, and a single experiment is insufficient to distinguish between all possibilities. In this case, multiple tasks may be necessary to estimate them unambiguously. For instance, this is why most experiments include control studies—to isolate certain parameters and reduce the number of parameters each experiment effectively estimates. This framework lets the experimenter simulate the experiment using ideal-observer-model predictions ahead of time to determine whether the experiment is sensitive and selective for distinguishing among individual hypotheses.

Limitations of Bayes' Nets and Statically Structured Generative Models

While humans incorporate vast contextual information to aid perception (Oliva & Torralba, 2007), Bayes' nets are best suited for representing situations in which several property variables are known to exist, but their state is uncertain. This limits the set of perceptual situations well characterized by Bayes' nets because the structures are predefined, and modifying them is not trivial. For instance, the generative process of a kitchen may not be well modeled by a Bayes' net because some objects may occur only infrequently (e.g., a waffle maker), may have widely varying sets of parts (e.g., overhead lamps can have very different numbers/types of bulbs), and exhibit unique hierarchical and recurrent patterns (e.g., a faucet is typically part of a sink but may instead be part of a refrigerator

door). Although theoretically Bayes' nets are able to model such generative processes, they are inefficient because many nodes will never take values. Moreover, an interesting structure that constrains the set of possible scenes is not explicitly or efficiently represented. Nonparametric methods have recently been applied in computer-vision applications to overcome this type of problem and aid visual inference (Sudderth & Jordan, 2009; Sudderth, Torralba, Freeman, & Willsky, 2008). These models use Bayes-net formalism to define abstract relationships among classes of objects and scenes, and nonparametric clusters to characterize specific instances of objects and scenes. This allows the models to share properties across similar objects and scenes while allowing specific instances to have unique and rich properties of their own.

Graph grammars are models that define rules for creating graph instances—in a sense they are a generative process for making generative processes—and may also be used to overcome **(p.60)** some of the limitations of Bayes' nets. For example, a graph grammar may specify that each object in a scene must have material and spatial properties, and they generate visual cues as long as they are not occluded by any other object. And a grammar may define how parts are shared across multiple object instances, such as similar cabinet doors/handles across different cabinets in a kitchen. Recently, probabilistic approaches to computer vision have begun achieving success by applying graph grammars to aid object recognition and image segmentation (Aycinena, Kaelbling, & Lozano-Perez, 2008; Han & Zhu, 2009; Zhu, Chen, & Yuille, 2009). In addition, graph grammars have been used to explain cognitive behaviors in a variety of situations (Kemp & Tenenbaum, 2008; Tenenbaum, Griffiths, & Niyogi, 2007).

CONCLUSION

Human perception entails a sophisticated reasoning strategy that combines sensory measurements with internal knowledge to construct accurate, detailed estimates about the state of the world. Generative knowledge is a useful formalism for characterizing observers' internal knowledge about the relationships among world properties and how they generate sensory cues. An observer can achieve Bayes-optimal perceptual inference if the generative knowledge accurately reflects the true generative process.

The challenges presented in the first section characterize the fundamental difficulties perceptual processing must overcome, and generative knowledge provides a natural solution to each challenge. Specifically, generative knowledge allows prior knowledge and indirect, auxiliary cues to disambiguate perception by ruling out unlikely and inconsistent potential interpretations. Generative knowledge can characterize relationships among world properties to allow vast contextual information to influence perception.

We presented a number of studies that *qualitatively* support discounting and explaining-away behavior in humans. An outstanding question is whether human perception is *quantitatively* consistent with Bayesian explaining away. By using the experimental framework illustrated by Figure 3.8 it will be possible to conduct strong quantitative tests of humans' generative knowledge.

REFERENCES

Bibliography references:

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*, 1057–1058.

Aycinena, M., Kaelbling, L. P., - Kersten, D. (2010). Within- and cross-modal distance information disambiguates visual size perception. *PLoS Computational Biology*, *6*(3), e1000697.

Battaglia, P. W., Schrater, P. R., & Kersten, D. (2005). Auxiliary object knowledge influences visually-guided interception behavior. *Proceedings of the 2nd Symposium on Applied Perception, Graphics, and Visualization, ACM International Conference Proceeding Series*, *95*, 145–152.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80.

Blake, R., Sobel, K. V., & James, T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science*, *15*, 397–402.

Blessing, W. W., Landauer, A. A., & Coltheart, M. (1967). The effect of false perspective cues on distance- and size-judgments: An examination of the invariance hypothesis. *The American Journal of Psychology*, *80*, 250–256.

Brenner, E., & van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research*, *39*, 975–986.

Epstein, W., Park, J., & Casey, A. (1961). The current status of the size-distance hypotheses. *Psychological Bulletin*, *58*, 491–514.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Gogel, W. C., Wist, E. R., & Harker, G. S. (1963). A test of the invariance of the ration of perceived size to perceived distance. *The American Journal of Psychology*, *76*, 537–553.

Granrud, C. E., Haake, R. J., & Yonas, A. (1985). Infants'sensitivity to familiar size: The effect of memory on spatial perception. *Perception and Psychophysics*, *37*, 459–466.

Gruber, H. E., & Dinnerstein, A. J. (1965). The role of knowledge in distance-perception. *The American Journal of Psychology*, *78*, 575–581.

Han, F., & Zhu, S. C. (2009). Bottom-up/top-down image parsing with attribute graph grammar. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 59–73.

Heinemann, E. G., & Nachmias, J. (1965). Accommodation as a cue to distance. *The American Journal of Psychology*, *78*, 139–142.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance

variant. *The American Journal of Psychology*, *54*, 21–37.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621–3629.

James, T. W., & Blake, R. (2004). Perceiving object motion using vision and touch. *Cognitive, Affective, and Behavioral Neuroscience*, *4*, 201–207.

Kaufman, L., & Rock, I. (1962). The Moon Illusion, I: Explanation of this phenomenon was sought through the use of artificial moons seen on the sky. *Science*, *136*, 953–961.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*, 718–727.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*, 10687–10692.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Knill, D. C. (1998a) Discriminating planar surface slant from texture: Human and ideal observers compared. *Vision Research*, *38*, 1683–1711.

Knill, D. C. (1998b). Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Research*, *38*, 1655–1682.

Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, *351*, 228–230.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation for perception and action. *Trends in Neuroscience*, *27*, 712–719.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*, 320–326.

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

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research*, *41*, 2653–2668.

Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, *82*, 76–89.

Mon-Williams, M., & Tresilian, J. R. (1999). The size-distance paradox is a cognitive phenomenon. *Experimental Brain Research*, *126*, 578–582.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of

cortico-cortical loops. *Biological Cybernetics*, *66*, 241–251.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*, 520–527.

Ono, H. (1966). Distal and proximal size under reduced and non-reduced viewing conditions. *The American Journal of Psychology*, *79*, 234–241.

Ono, H., Muter, P., & Mitson, L. (1974). Size-distance paradox with accommodative micropsia. *Perception and Psychophysics*, *15*, 301–307.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufmann. Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Strat, T. M., & Fischler, M. A. (1991). Context-based vision: Recognizing objects using information from both 2-D and 3-D imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 1050–1065.

Sudderth, E., & Jordan, M. I. (2009). Shared segmentation of natural scenes using dependent Pitman-Yor processes. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21* (pp. 1585–1592).

Sudderth, E., Torralba, A., Freeman, W. T., & Willsky, A. (2008). Describing visual scenes using transformed objects and parts. *International Journal of Computer Vision*, *1–3*, 291–330.

Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 301–322). Oxford, England: Oxford University Press.

Yonas, A., Granrud, C. E., & Pettersen, L. (1985). Infants'sensitivity to relative size information for distance. *Developmental Psychology*, *21*, 161–167.

Yonas, A., Pettersen, L., & Granrud, C. E. (1982). Infants'sensitivity to familiar size as information for distance. *Child Development*, *53*, 1285–1290.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Science*, *10*, 301–308.

Zhu, L., Chen, Y., & Yuille, A. (2009). Unsupervised learning of probabilistic grammar-Markov models for object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 114–128.

Notes:

(1) The term *belief* is used in a statistical sense, referring to information held by the

observer about the external world state.