Object Perception: Generative Image Models and Bayesian Inference

Daniel Kersten[1]

Psychology Department, University of Minnesota
75 East River Road,
Minneapolis, Minnesota, 55455
U.S.A.
kersten@umn.edu
http://kersten.org

Abstract. Humans perceive object properties such as shape and material quickly and reliably despite the complexity and objective ambiguities of natural images. The visual system does this by integrating prior object knowledge with critical image features appropriate for each of a discrete number of tasks. Bayesian decision theory provides a prescription for the optimal utilization of knowledge for a task that can guide the possibly sub-optimal models of human vision. However, formulating optimal theories for realistic vision problems is a non-trivial problem, and we can gain insight into visual inference by first characterizing the causal structure of image features–the generative model. I describe some experimental results that apply generative models and Bayesian decision theory to investigate human object perception.

1 Object surface interactions

Consider a collection of objects in a scene. Given an image of their surfaces, one can ask many questions: Do the surfaces belong to the same object? If part of the same object, how are they oriented with respect to each other? If separate, is one occluding the other? Are they in contact? How far apart? What kind of materials are they made of? What color? Answers to each of these questions requires the definition of a visual task. Task definition declares some variables more useful than others, and thus which need to be made explicit and accurately estimated. When the visual system answers these questions, it has solved a complex inference problem. We better understand the nature of visual ambiguity and its resolution by first considering how image features are generated through the combination and interaction of potentially useful scene variables (e.g. object shape) with other scene variables that may be less useful (e.g. illumination direction). Generative models help to identify the key information used by human visual perception, and thus provide a basis for modeling vision as Bayesian statistical inference [27, 16, 34].

Modeling the image formation or generative process makes explicit the causal structure of image features. Identifying causal influences on the image is typically

well-defined, and thus easier than the inverse problem of inference. A generative model helps to make clear where the ambiguities lie, and set the stage for psychophysical inquiry into what variables are important to human vision, as well as to identify and simplify the constraints needed to solve the computational inverse problem [27]. Generative models describe the probability of an image description I, as a function of key causal factors in the scene S. Both knowledge of image formation, $p(I|S)$, and prior knowledge $p(S)$ contribute to the generative model. Such a model can be either image-based or scene-based (cf. [35] and [13]). Image-based models seek concise statistical descriptions of an image ensemble (e.g. all images of apples). Examples include texture models [32] and 2D shape models in terms of deformable templates [10]. Scene-based models describe image ensembles in terms of scene constructions, using computer graphics [6]. In either case, a generative model identifies the factors that characterize image variability, making it possible to experimentally test which ones are important for a human visual task. We describe experimental results from several scene-based models in the examples below.

I will next provide an overview of vision as statistical inference, focusing on three classes of problems: invariance, cue integration, and perceptual "explaining away". Then I will illustrate each of these with psychophysical results on: 1) Perception of depth and color given illumination variation; 2) Perception of surface contact; and 3) Perceptual organization given occlusion. Finally, I address the question of whether the visual brain may recapitulate aspects of the generative model in order to test its own models of incoming visual measurements.

2   Invariance, Cue integration & "Explaining away"

From a Bayesian perspective, knowledge is specified in terms of a joint probability distribution on all relevant variables, both image measurements and object variables. It is helpful to characterize object inference problems in terms of a graph that illustrates how image measurements are influenced by the object hypotheses [29, 30]. Many object perception studies fall into one of three simple sub-cases which will here be referred to as invariance, cue integration, and "explaining away" (Figure 1). The generative model expressed by the graph can be interpreted as specifying how the joint probability is factored into the conditional probabilities[1].

---

[1] If I_i and S_j indicate the ith and j^{th} image and object variables respectively, then $p(..., I_i..., S_j, ...)$ is the joint probability. Invariance, cue integration and the explaining away example have joints: $p(I_1, S_1, S_2)$, $p(I_1, I_2, S_1)$ and $P(I_1, I_2, S_1, S_2)$. The influence relations simplify the joint probability distributions:

$$p(I, S_1, S_2) = p(I|S_1, S_2)p(S_1)p(S_2)$$

$$p(I_1, I_2, S_1) = p(I_1, I_2|S_1)p(S_1) \ and$$

$$p(I_1, I_2, S_1, S_2) = p(I_2|S_2)p(I_1|S_1, S_2)p(S_1)p(S_2)$$

.

The task definition adds additional constraints to the estimation problem in specifying which nodes are fixed measurements (black), which are variables to be estimated (green), and which are confounding variables to be discounted (red; See Figure 1). Discounting can be formalized with a utility function (or its complement, a loss function). Visual ambiguity is often reduced by auxiliary measurements (yellow node) that may be available in a given image, or actively sought. These auxiliary measurements may provide diagnostic information regarding a confounding variable, and as a consequence help to explain away ambiguity in another image measurement that pertains more directly to the useful target variable of interest. "Explaining away" refers to the case when new or auxiliary evidence under-cuts an earlier explanation [26].
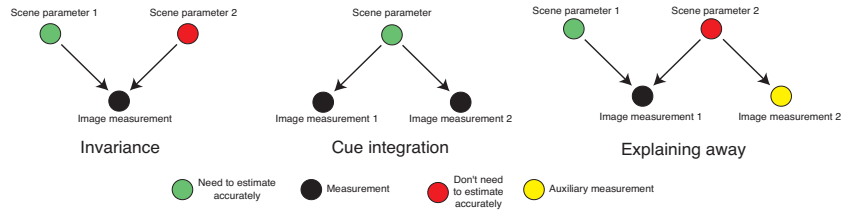

Fig. 1. Graphs for three classes of generative models. The nodes represent random variables that fall into four classes. The variables may be: 1) known (black); 2) unknown and need to be estimated accurately (green); 3) unknown, but do not need to be explicitly and accurately estimated (red); 4) not directly influenced by the object variable of interest, but may be useful for resolving ambiguity (yellow). The arrows indicate how scene or object properties influence image measurements or features. Left panel illustrates a causal structure that gives rise to the invariance problem. (See Section 2.1.) Middle panel illustrates cue integration (See Section 2.2). Right panel illustrates a case that can lead to "explaining away" (See Sections 2.3 and 3).

Bayesian decision theory combines probability with task requirements to derive quantitative, optimal theories of perceptual inference [15, 8, 3]. Human perception is often surprisingly close to optimal, thus such "ideal observer" theories provide a good starting point for models of human vision [8]. The basic concepts are illustrated with a simple example in Figure 2. A flat elliptical object in 3D projects an ellipse onto the image. One can measure the aspect ratio of the image of the ellipse. This information constrains, but does not uniquely determine the aspect ratio of the ellipse in the three-dimensional world. A unique estimate can be made that depends on combining prior knowledge and task utility assumptions. A special case assumes there is a uniform cost to all errors in the estimates of the confounding variable (the green bar in Figure 2 would span the whole space in one direction). This case corresponds to marginalizing or integrating out the confounding variable from the posterior probability $p(S_1, S_2|I_1)$. So for example, inference in the invariance case requires finding the value of $s$cene

parameter 1 (S_1) that maximizes: $\int_{S_2} p(I_1|S_1, S_2)p(S_1)p(S_2)/p(I_1)dS_2$, where scene parameter 2 (S_2) is the confounding variable, and I_1 is the image feature.

2.1   Invariance: Discounting confounding variables

How does the visual system enable us to infer the same object despite considerable image variation due to viewpoint, illumination, occlusion, and background changes? This is the well-known problem of invariance, or object constancy. Here the target variable is constant, but the image measurements vary as a function of variations in the confounding variable (Left panel of Figure 1). Confounding variables play the role of "noise" in classical signal detection theory; however, the generative modeling is typically more complex (as illustrated by 3D computer graphics synthesis), and the formal inference problem can be complex involving high dimensions, non-Gaussian distributions, and non-linear estimators. Illumination variation is perhaps the most dominant source of variation for the tasks of vision. Let's look at illumination variation in the context of two tasks, depth and material perception.

Illumination variation.   The vagaries of illumination generate enormous variations in the images of an object. Typically illumination results from an infinite number of point sources, both direct (luminous) and indirect (reflected). Illumination varies in dominant direction, level, spatio-temporal distribution, and spectral content. Further, it interacts with surface properties to produce complex effects of specular reflection. These are confounding variables for many of the tasks of object perception.

How far apart are two objects?. Cast shadows are an effective cue for relative depth [22], despite ambiguity between relative depth between the casting object and the background, and light source angle. At first one might guess that the visual system requires accurate knowledge of the lighting arrangement in order to estimate depth from shadows. However, if one assumes that there is uniform cost to errors in light source slant estimates, decision theory analysis can predict, based on the geometry alone that cast shadows should be most reliable when near the object, and that the "optimal" estimate of object location is that it is as far from the background as the shadow is from the object [14].

What is the material color of an object?. Color constancy has been studied for well over a century. Everyday variations in the spectral content, levels, and gradients of the illuminant have relatively little effect on our perception of surface color. One way of formalizing the problem is to understand how the objective surface invariant, surface reflectivity, can be estimated given variations in illumination [3]. Most such studies have been restricted to the perception of material color or lightness on flat surfaces with no illumination contributions from neighboring surfaces. Bloj, Kersten, and Hurlbert [1] showed that color perception is influenced by the 3D arrangement of a nearby surface. They constructed a chromatic version of the classic Mach Card (Figure 3). With it, they showed that a white surface appears white when its pinkish tinge can be explained in terms of a near facing red surface, but appears pigmented pink when the red surface appears to be facing away. The experimental results showed that the
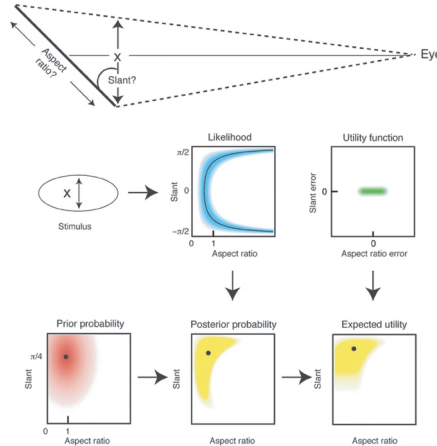
Fig. 2. Example of applying Bayesian theory to the problem of estimating the slant α and *aspect ratio d in 3D* of a flat ellipse, given *x the aspect ratio measured in the image*. The generative model $x = dsin(\alpha) + noise$ is well-defined and tells us how scene variables determine an image measurement x. Generative knowledge determines the likelihood function $p(x|\alpha, d)$. The Bayesian observer first computes the likelihood of stimulus x for each pair of scene values α, d. The solid black curves in the likelihood plot show the combinations of slant and aspect ratio that are exactly consistent with with the generative model if there were no noise. Non-zero likelihoods occur because of noise in the measurement of x. The Bayesian observer then multiplies the likelihood function by the prior probability distribution for each pair of scene values to obtain the posterior probability distribution, $p(\alpha, d|x)$. The prior probability distribution corresponds to the assumption that surface patches tend to be slanted away at the top and have aspect ratios closer to 1.0. Accuracy along some dimensions can be more important than along other dimensions depending on the task. For example, recognizing a a particular tea-cup could require accurate estimation of aspect ratio of the top, but not the slant with respect to the viewpoint. In this case slant is the confounding variable. On the other hand, stepping on to a flat stone requires accurate estimation of the slant, but not the aspect ratio. Thus, the 3D aspect ratio is the confounding variable. To take task-dependence into account, the posterior probability distribution is convolved with a utility function, representing the costs and benefits of degrees of accuracy, to obtain the expected utility associated with each interpretation. The Bayesian decision theory observer picks the interpretation that maximizes the expected utility, as indicated by the black dot in the lower right panel. (Black dots and curves indicate the maximum values in the plots.) The asymmetric utility function would correspond to the assumption that it is more important to have an accurate estimate of slant than aspect ratio. Figure reprinted with permission ?? from Nature Neuroscience.

human visual system has intrinsic knowledge of mutual illumination or inter-reflections–i.e. how the color of light from near-by surfaces can confound image measurements. A Bayesian ideal observer that has generative knowledge of indirect lighting and that integrates out contributions from illumination direction predicted the central features of the psychophysical data and demonstrated that this shape-color contingency arises because the visual system "understands" the effects of mutual illumination [1, 7] [2].

2.2   Cue Integration

Cue integration is a well-known problem in perceptual psychology. For example, one can identify over a dozen cues that the human visual system utilizes for depth perception. In computer vision and signal processing, cue integration is studied under the more general rubric of "sensor fusion" [2]. There has been recent interest in the degree to which the human visual system combines image measurements optimally. For example, given two conflicting cues to depth, the visual system might get by with a simple averaging of each estimate, even though inaccurate. Or it may determine that one measurement is an outlier, and should not be integrated with the other measurement [17, 4]. The visual system could be more sophisticated and combine image measurements weighted according to their reliability [12, 33]. These issues have their roots in classical questions of information integration for signal detectability, e.g. probability vs.

---

[2] The Bayesian calculation goes as follows. The target variable of interest is the reflectivity ($S_1 = \rho$) (measured in units of chroma). The likelihood is determined by either a one-bounce (corner) or zero-bounce generative model (roof condition) of illumination. Assume that the shape is fixed by the stereo disparity, i.e. condition on shape (*roof or corner*). From the one-bounce model, the intensity equation for white pigmented side (surface 1) is:

$$I_1(\lambda, x, \rho, E, \alpha_1, \alpha_2) = E(\lambda)\rho_1 * (\lambda)[cos\alpha_1 + f_{21}\rho_2(\lambda)cos\alpha_2]$$

where the first term represents the direct illumination with respect to the surface and the second term represents indirect illumination due to light reflected from the red side (surface 2) [5]. $f_{21}(x)$ is the form factor describing the extent to which surface 2 reflects light onto surface 1 at distance x from the vertex [6]. The angles α_1 and α_2 denote the angle between the surface normal and the light source direction for surfaces 1 and 2 respectively. $E(\lambda)$ is the irradiance as a function of wavelength λ For the zero-bounce generative model (roof condition), the form factor $f_{21} = 0$, so that:

$$I_1(\lambda, x, \rho, E, \alpha) = E(\lambda)\rho_1 * (\lambda)cos\alpha_1$$

These generative models determine the likelihood functions. Observers do not directly measure I_1, but rather chroma C_{obs}, modeled by the capture of light by the retinal cones. When an observer is asked to match the surface color to the ith test patch, the optimal decision is based on $P(\rho_1^i|C_{obs})$, which is obtained by integrating out x and the confounding variable α from $p(C_{obs}|\rho^i, x, \alpha, E)$. To a first approximation, observers' matches were predicted well by an observer which is ideal apart from an internal matching variability. For more details see [1, 31].
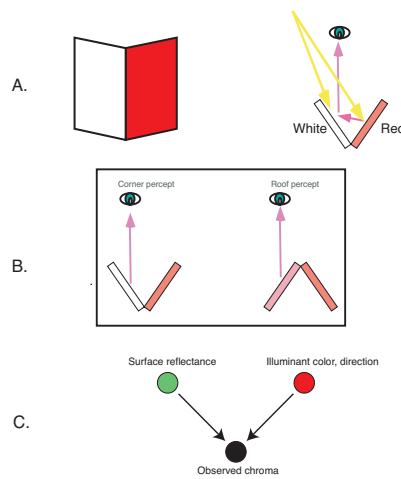
Fig. 3. A. The "colored Mach card" consists of a white and red half [1]. It is folded such that the sides face each other. The viewer's task is to determine the material color of the white side, given the viewing and illumination arrangement illustrated. B. If the card's shape is seen as it truly is (a concave "corner"), the white side is seen as a white card, tinted slightly pink from the reflected red light. However, if the shape of the card *appears* as though the sides face away from each other (convex or "roof" condition), the white card appears pink–i.e. more saturated towards the red. Note that there may be little or no difference in the image information for these two percepts. C. The black, green and red nodes represent an image measurement (e.g. pinkishness), a scene hypothesis (is the material's spectral reflectivity closer to that of white or pink pigmented paper?), and a confounding variable (illumination direction), respectively. See Section 2.1).

information summation [9]. Even when we do not have a specific idea of what image information vision uses when integrating cues, we can sometimes investigate the information in terms of the scene variables that contribute. So while a quantitative description of the relevant image measurements may be lacking, this approach has the advantage of using realistic images. Further, even without an objective formulation of the problem and its ideal observer, psychophysics can provide insights into how well cues are integrated. This is illustrated in the following example.

Are two surfaces in contact?. Surface contact decisions are a special case of relative depth estimation, whose effects in the image are the result of surface and illumination interactions as discussed earlier in Section 2.1. Determining whether or not two surfaces are in contact is a common visual function, useful for deciding whether the surfaces belong to an object, or if an object is detachable or graspable. What is the visual information for the perception of surface contact? The interaction of light with surfaces in close proximity results in characteristic shadows as well as in surface inter-reflections. Inter-reflections and shadows can each potentially provide information about object contact (Figure 4). Psychophysical measurements of contact judgments show that human observers combine image information from shadows with inter-reflections to achieve higher sensitivity than when only shadows or inter-reflections are present [21].

2.3   "Explaining Away"

Different scene variables can give rise to different kinds of image measurements. Conversely, different image measurements in the same, or subsequently acquired images (e.g. fixations), can be differentially diagnostic regarding their causes in terms of object properties. The generative model can provide insights into what information should, in principle, help to disambiguate hypotheses regarding the properties of a target object.

Object color, shape and mutual illumination revisited. We illustrate perceptual "explaining away" by revisiting the colored Mach card of Figure 3. Because of the ambiguity of perspective, a rigid folded card can appear as concave or convex from a fixed viewpoint. Stereo disparity can provide reliable information for one or the other shape interpretations. When this happens, the shape hypothesis changes with the surface material hypothesis in explaining the pinkish tinge of the observed white pigmented card face as shown in Figure 5. Examples of this type of inference occur more generally in the context of Bayes networks [29]. The relevant concepts are also related to the idea of "strong fusion" [2].

There are many examples where explaining away does not work in human perception, and we may ultimately gain more insight into the mechanisms of vision from these cases. Mamassian et al. (1998) describe an example where a pencil that casts a shadow over a folded card fails to disambiguate the shape of the card, resulting in physically inconsistent perceptions of the shadow and geometry [22].
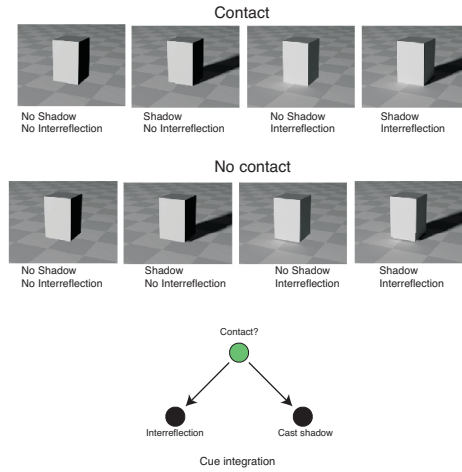
Fig. 4. A. Computer generated images of a box on an extended textured ground plane that was either in contact with the ground plane or slightly above it [21]. Images were rendered for four conditions: 1) no shadow plus no inter-reflection, 2) shadow only, 3) inter-reflection only, and 4) shadow plus inter-reflection. Observers were required to judge the degree of contact for each image. In the images with no shadow or inter-reflections, observers performed at chance. Inter-reflections, shadows, and a combination of inter-reflections and shadows all resulted in a high sensitivity for judging object contact. Information from shadows and inter-reflections was combined to result in near-perfect judgement of surface contact. B. The graphical structure for the cue integration problem. The green node represents an hypothesis of contact or not, and the black nodes image measurements or evidence (i.e. the image effects of a cast shadow and/or mutual illumination. Figure adapted with permission ?? from Perception & Psychophysics.
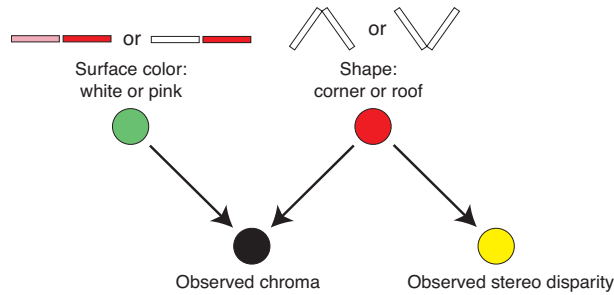

Fig. 5. Illustrates "explaining away" (Sections 2.3 and 3). One hypothesis (pink paint) may explain a "pinkish" image chroma measurement, but another hypothesis (nearby red surface) could also explain the pinkish chroma, but in terms of indirect reddish illumination. An auxiliary image measurement (yellow node, disparity indicating a concave relationship between a white and red surface) could tip the balance, and the joint hypothesis "concave white-red card" could explain both image measurements with high probability. The pink pigment hypothesis is no longer probable.

3   Perceptual "explaining away" in the Brain?

The primate visual system is composed of a hierarchy of more than thirty visual areas, pairs of which communicate through both feedforward and feedback connections. A possible role for higher-level visual areas may be to represent hypotheses regarding object properties that could be used to resolve ambiguities in the incoming retinal image measurements. These hypotheses could predict incoming data through feedback and be tested by computing a difference signal or residual at the earlier level [24, 28]. Thus, low activity at an early level would mean a "good fit" or explanation of the image measurements. One way of testing this idea is to use fMRI to compare the activity of early areas for good and bad fits given the same incoming retinal signal.

Consider the problem of perceiving a moving occluded diamond as shown in Figure 6A. The four moving line segments can appear to cohere as parts of a single horizontally translating diamond or can appear to have separate vertical motions [19]. In order to perceive a moving object as a whole, the brain must measure local image features (motion direction and speed), select those likely to belong to the same object, and integrate these measurements to resolve the local ambiguity in velocity [33]. The selection process involves choosing which of the four segments belong together, and this in turn is closely tied to the visual system accounting for the missing vertices as due to occlusion. Neurons in primary visual area V1 have spatially localized receptive fields selective for edge orientation and motion direction. A good fit to incoming data would occur when all four contour segments are perceptually grouped as a diamond. This happens when the segments appear to move horizontally in synchrony for the horizontally moving diamond percept. However, when the line segments appear to move separately or have other multiple groupings, the apparent movement of the segments not grouped is poorly predicted, resulting in a poorer fit to the local measurements in V1.

Experiments showed that when observers viewed the bistable stimulus of Figure 6A, fMRI BOLD activity in V1 *decreased* when the segments were perceived to be part of a single object. Further, in other experiments, the BOLD response to visual elements that appeared either to be grouped into objects or incoherently arranged showed reductions of activity in V1 when elements formed coherent shapes (See Figure 6; [25]). One might further conjecture that activation in higher-level areas should show the opposite direction of activity change. The lateral occipital complex (LOC) is a higher level object processing area that has received considerable recent attention [18, 11]. Measurements here showed increases in LOC activity were concurrent with reductions of activity in primary visual cortex (V1) when elements formed coherent shapes [25]. These results are consistent with the idea that activity in early visual areas may be reduced as a result of object hypotheses represented in higher areas.

In a general sense, feedback between visual areas may be the internal recapitulation of the external generative processes that give rise to the images received.
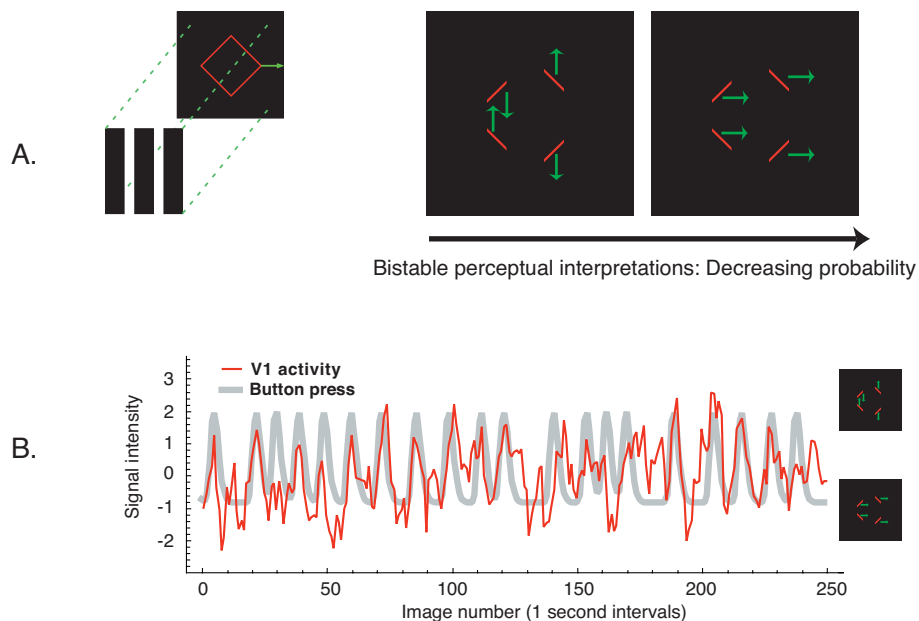
Bistable perceptual interpretations: Decreasing probability

Fig. 6. A. Occluded view of a translating diamond generates ambiguous perceptual interpretations. The diamond's vertices are covered by black occluders so that an observer sees just four moving line segments [19, 20]. The perception is bistable: the four segments appear to be moving in different vertical directions, or to cohere as part of a horizontally moving diamond. B. fMRI activity in human V1 (red) predicts an observer's reported perceptual state (thick gray lines). Further, fMRI activity *de*creases when the segments are perceived to be part of a single object. A similar pattern of results is found for other manipulations [25]. These findings are consistent with predictive coding models of vision in which the inferences of higher-level visual areas inhibit incoming sensory signals in earlier areas through cortical feedback.

References

1. Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception via mutual illumination. *Nature*, **402**, 877-879.
2. Clark, J. J., & Yuille, A. L. (1990). **Data Fusion for Sensory Information Processing**. Boston: Kluwer Academic Publishers.
3. Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *J Opt Soc Am A*, **14**, (7), 1393-411.
4. Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America, A*, **5**, (10), 1749-1758.
5. Drew, M., & Funt, B. (1990). Calculating surface reflectance using a single-bounce model of mutual reflection. *Proceedings of the 3rd International Conference on Computer Vision* Osaka: 393-399.
6. Foley, J., van Dam, A., Feiner, S., & Hughes, J. (1990). **Computer Graphics Principles and Practice**, (2nd ed.). Reading, Massachusetts: Addison-Wesley Publishing Company.
7. Gegenfurtner, K. R. (1999). Reflections on colour constancy. *Nature*, **402**, 855-856.
8. Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nat Neurosci*, **5**, (6), 508-10.
9. Green, D. M., & Swets, J. A. (1974). **Signal Detection Theory and Psychophysics**. Huntington, New York: Robert E. Krieger Publishing Company. 1974.
10. Grenander, U. (1996). **Elements of Pattern theory**. Baltimore: Johns Hopkins University Press.
11. Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res*, **41**, (10-11), 1409-22.
12. Jacobs, R. A. (2002). "What determines visual cue reliability?" Trends Cogn Sci 6(8): 345-350.
13. Kersten, D. (1997). Inverse 3D Graphics: A Metaphor for Visual Perception. *Behavior Research Methods, Instruments, & Computers*, **29**, (1), 37-46.
14. Kersten, D. (1999). High-level vision as statistical inference. In Gazzaniga, M. S. (Ed.), *The New Cognitive Neurosciences – 2nd Edition*(pp. 353-363). Cambridge, MA: MIT Press.
15. Kersten, D., & Schrater, P. R. (2002). Pattern Inference Theory: A Probabilistic Approach to Vision. In Mausfeld, R.,& Heyer, D. (Ed.), *Perception and the Physical World* (pp. Chichester: John Wiley& Sons, Ltd.
16. Knill, D. C., & Richards, W. (1996). **Perception as Bayesian Inference**. Cambridge: Cambridge University Press.
17. Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.
18. Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion Effects in the Human Lateral Occipital Complex. *Cereb Cortex*, **12**, (2), 163-77.
19. Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Res*, **32**, (2), 263-73.
20. Lorenceau, J., & Alais, D. (2001). Form constraints in motion binding. *Nat Neurosci*, **4**, (7), 745-51.
21. Madison, C., Thompson, W., Kersten, D., Shirley, P., & Smits, B. (2001). Use of interreflection and shadow for surface contact. *Perception and Psychophysics*, **63**, (2), 187-194.

22. Mamassian, P., Knill, D. C., & Kersten, D. (1998). The Perception of Cast Shadows. *Trends in Cognitive Sciences*, **2**, (8), 288-295.

23. McDermott, J., Weiss, Y., & Adelson, E. H. (2001). Beyond junctions: nonlocal form constraints on motion interpretation. *Perception*, **30**, (8), 905-23.

24. Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern*, **66**, (3), 241-51.

25. Murray, S. O., Kersten, D, Olshausen, B. A., Schrater P., & Woods, D.L. (Under review) Shape perception reduces activity in human primary visual cortex. Submitted to the *Proceedings of the National Academy of Sciences.*

26. Pearl, J. (1988). **Probabilistic reasoning in intelligent systems : networks of plausible inference**, (Rev. 2nd printing. ed.). San Mateo, Calif.: Morgan Kaufmann Publishers.

27. Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, **317**, 314-319.

28. Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects [see comments]. *Nat Neurosci*, **2**, (1), 79-87.

29. Ripley, B. . **Pattern Recognition and Neural Networks**. Cambridge University Press. 1996.

30. Schrater, P. R., & Kersten, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision*, **40**, (1), 73-91.

31. Schrater, P., & Kersten, D. (2001). Vision, Psychophysics, and Bayes. In Rao, R. P. N., Olshausen, B. A., & Lewicki, M. S. (Ed.), *Probabilistic Models of the Brain: Perception and Neural Function*(pp. Cambridge, Massachusetts: MIT Press.

32. Simoncelli, E. P. (1997). Statistical Models for Images: Compression, Restoration and Synthesis. Pacific Grove, CA.: IEEE Signal Processing Society.

33. Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, **5**, (6), 598-604.

34. Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D.C., K., & W., R. (Ed.), *Perception as Bayesian Inference*(pp. Cambridge, U.K.: Cambridge University Press.

35. Zhu, S.C., Wu, Y., and Mumford, D. (1997). "Minimax Entropy Principle and Its Application to Texture Modeling". Neural Computation. **9**(8).