Computational Vision
U. Minn. Psy 5036
Daniel Kersten
Lecture 23: Perceptual integration & cooperative computation

Initialize

Spell check off

```
Off[General::spell1];
```

Outline

Last time

Scientific writing

The importance of logical flow/transitions. Throughout, set the context to create an expectation, which is subsequently fulfilled. This applies across all units, sentence to sentence, paragraph to paragraph, and section to section. Minimize jargon and unnecessary words. It can help to keep sentences short and direct.

Introduction: Motivate by describing what is known, what is unknown, and why it is important to know. Explain the question(s) you will answer to advance what is known.
Methods: Describe how data were collected, or simulations made, and the methods of analysis. Provide sufficient detail to enable replication.
Results: Your results should parallel the closing questions of the Introduction. It usually helps to work on the figures first before writing.
Discussion: Summarize the main results, but not at the expense of being overly redundant. Describe the implications for the field, and if possible propose further questions to be addressed.

Today

Integrating perceptual information

Modular vs. cooperative computation

To make problems tractable, most theories of visual estimation have been “modular”, e.g. surface-color-from-radiance (Land, 1959), shape-from-shading (Horn, 1975), optic flow (Hildreth, 1983) or structure-from-motion (Ullman, 1979). While there is evidence for multiple pathways and areas in the brain, with possible analogous functions, we have only sketchy ideas of their computations, and the extent to which the computational modules of theorists may relate to cortical architecture.

In this lecture, we focus on the problem of how information is integrated. It is phenomenally apparent that visual information is integrated to provide a strikingly singular description of the visual environment. By looking at how human perception integrates cues, and scene attributes, we will get some idea of how different kinds of visual processing in the brain might interact, and what kind of information is represented.

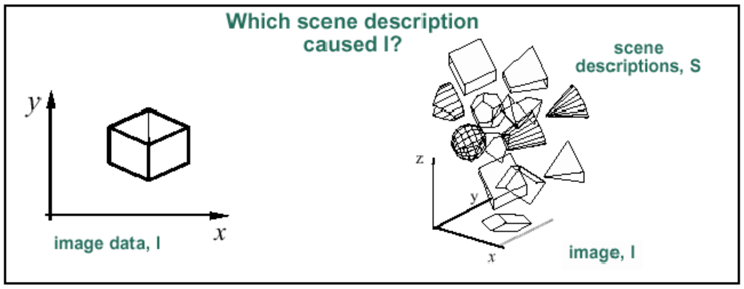
Some basic graph types in vision (Review)

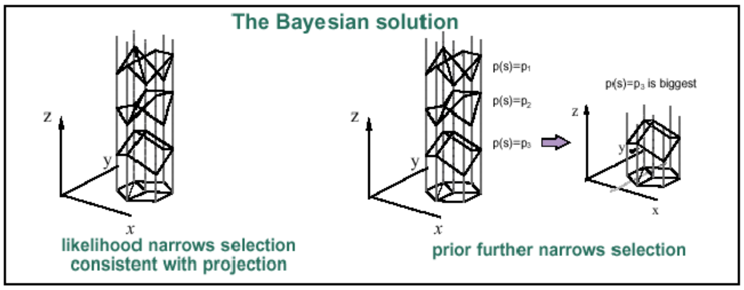
See: Kersten, D., & Yuille, A. (2003) and Kersten, Mamassian & Yuille (2004)

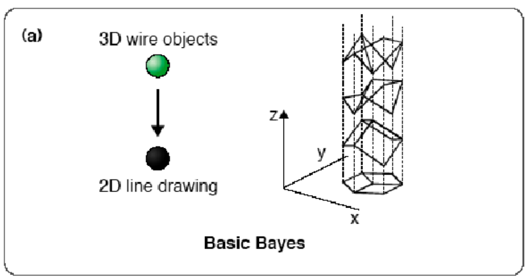
Basic Bayes

$$p[S \mid I] = \frac{p[I \mid S]\, p[S]}{p[I]}$$

Usually, we will be thinking of the **Y** term as a random variable over the hypothesis space, and **X** as data. So for visual inference, **Y = S** (the scene), and **X = I** (the image data), and **I = f(S)**.
We'd like to have:
p(S|I) is the **posterior** probability of the scene given the image
-- i.e. what you get when you condition the joint by the image data. The posterior is often what we'd like to base our decisions on, because as we discuss below, picking the hypothesis **S** which maximizes the posterior (i.e. maximum a posteriori or **MAP** estimation) minimizes the average probability of error.
p(S) is the **prior** probability of the scene.
p(I|S) is the **likelihood** of the scene. Note this is a probability of **I**, but not of **S**.







We've seen that the idea of prior assumptions that constrain otherwise underconstrained vision problems is a theme that pervades much of visual perception. Where do the priors come from? Some may be built in early on or hardwired from birth, and others learned in adulthood. See: Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. Nat Neurosci, 7(10), 1057-1058 for an example of learning the “light from above” prior for shape perception.

Low-level vision

We've seen a number of applications of Basic Bayes, including the algorithms for shape from shading and optic flow.

In 1985, Poggio, Torre and Koch showed that solutions to many of computational problems of low vision could be formulated in terms of maximum a posteriori estimates of scene attributes if the generative model could be described as a matrix multiplication, where the image I is matrix mapping of a scene vector S:

$$I = \mathbf{A}S$$

$$E = (I - \mathbf{A}S)^T (I - \mathbf{A}S) + \lambda S^T \mathbf{B} S$$

Then a solution corresponded to minimizing a cost function E, that simultaneously tries to minimize the cost due to reconstructing the image from the current hypothesis S, and a prior "smoothness"

constraint on S. λ is a (often free) parameter that determines the balance between the two terms. If there is reason to trust the data, then λ is small; but if the data is unreliable, then more emphasis should be placed on the prior, thus λ should be bigger.
For example, S could correspond to representations of shape, stereo, edges, or motion field, and smoothness be modeled in terms of nth order derivatives, approximated by finite differences in matrix B.

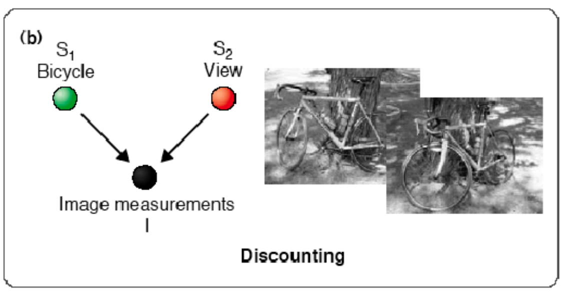
The Bayesian interpretation comes from multivariate gaussian assumptions on the generative model:

$$p(I \mid S) = k \times \exp\left[-\frac{1}{2\sigma_n^2}(I - \mathbf{A}S)^T (I - \mathbf{A}S)\right]$$

$$p(S) = k' \times \exp\left[-\frac{1}{2\sigma_s^2} S^T \mathbf{B} S\right]$$

However, most vision problems are not characterized by a linear generative model.

Discounting



This Bayes influence graph describes the case where the joint distribution can be factored as:
$p(s_1, s_2, I) = p(I|s_1,s_2)p(s_1)p(s_2)$

Optimal inference for this task requires that we calculate the marginal posterior:
$p(s_1|I) \propto \int_{S_2} p(s_1, s_2 \mid I)\, d s_2$

Liu, Knill & Kersten (1995) describe an example with:
I -> 2D x-y image measurements, s_1-> 3D object shape, and s_2-> view

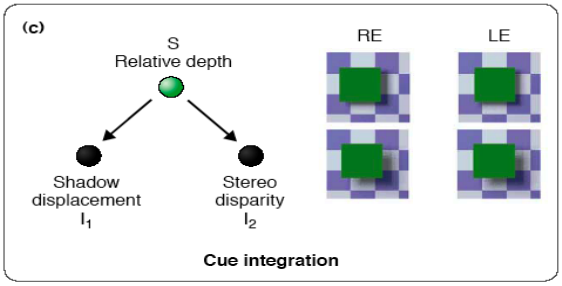
Bloj et al. (1999) have an example estimating s_1-> surface chroma (saturation) with s_2-> illuminant direction.
In the next lecture, we'll see a computationally tractable ideal observer analysis of object recognition given view variation.

Cue integration

Cue integration is a form of perceptual integration where low-level visual measurements are combined to produce a more accurate and reliable output.

Weak fusion

Clark & Yuille, Landy & Maloney, Schrater & Kersten.



This Bayes generative graph describes the factorization:

p(S,I1,I2) = p(I1|S)p(I2|S)p(S)

One consequence of this graph, is that one can show that the optimal combined estimate is the weighted sum of the separate estimates, where the weights w_i are determined by the relative reliabilities :

$$\hat{\mu}_{\text{combined}} = \hat{\mu}_{\text{cue1}} w_1 + \hat{\mu}_{\text{cue2}} w_2 = \hat{\mu}_{\text{cue1}} \frac{r_1}{r_1 + r_2} + \hat{\mu}_{\text{cue2}} \frac{r_2}{r_1 + r_2}.$$

This is a simple but important idea which raises the empirical question of whether human perception integrates cues optimally. We've seen this principle applied before when we studied Weiss et al.'s solution to the aperture problem in motion. Let's see how to derive it.

Maximum a posteriori observer for cue integration: conditionally independent cues

We'll change notation, and let x_1 and x_2 be image measurements, i.e. the cues. The simple Bayes graph shown above describes the case where the two cues are conditionally independent. We can think of s as an underlying cause of the two measurements. Assuming s is known, then p(x1,x2|s) = p(x1|s)p(x2|s).

Let's consider the simple Gaussian case where $x_i = \mu_{\text{cue}\,i} + n_i$. We can show that optimal combined cue estimate is a weighted average of the cues.

p(s|x1,x2) = p(x1,x2|s)p(s)/p(x1,x2) ∝ p(x1|s)p(x2|s) $= e^{-(x_1-s)^2/2\sigma_1{}^2} e^{-(x_2-s)^2/2\sigma_2{}^2}$

In[97]:= $\textbf{PowerExpand}\left[\textbf{Log}\left[E^{-(x_1-\mu)^2/(2\sigma_1{}^2)}\, E^{-(x_2-\mu)^2/(2\sigma_2{}^2)}\right]\right]$

Out[97]= $-\frac{(-\mu + x_1)^2}{2\sigma_1^2} - \frac{(-\mu + x_2)^2}{2\sigma_2^2}$

In[98]:= $D\left[-\frac{(x_1-\mu)^2}{2\,\sigma_1^2}-\frac{(x_2-\mu)^2}{2\,\sigma_2^2},\,\mu\right]$

Out[98]= $\frac{-\mu+x_1}{\sigma_1^2}+\frac{-\mu+x_2}{\sigma_2^2}$

In[99]:= $\text{Solve}\left[\frac{x_1-\mu}{\sigma_1^2}+\frac{x_2-\mu}{\sigma_2^2}==0,\,\mu\right]$

Out[99]= $\left\{\left\{\mu\to\frac{x_2\,\sigma_1^2+x_1\,\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right\}\right\}$

In[100]:= $\left\{\left\{\mu\to\frac{x_2\,\sigma_1^2+x_1\,\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right\}\right\}\,/.\,\left\{\sigma_1^2\to 1/r_1,\,\sigma_2^2\to 1/r_2\right\}$

Out[100]= $\left\{\left\{\mu\to\frac{\frac{x_1}{r_2}+\frac{x_2}{r_1}}{\frac{1}{r_1}+\frac{1}{r_2}}\right\}\right\}$

where $r_i\left(=\frac{1}{\sigma_i{}^2}\right)$, is called the reliability.

In[101]:= $\mu\to\frac{r_1\,x_1}{r_1+r_2}+\frac{r_2\,x_2}{r_1+r_2}$

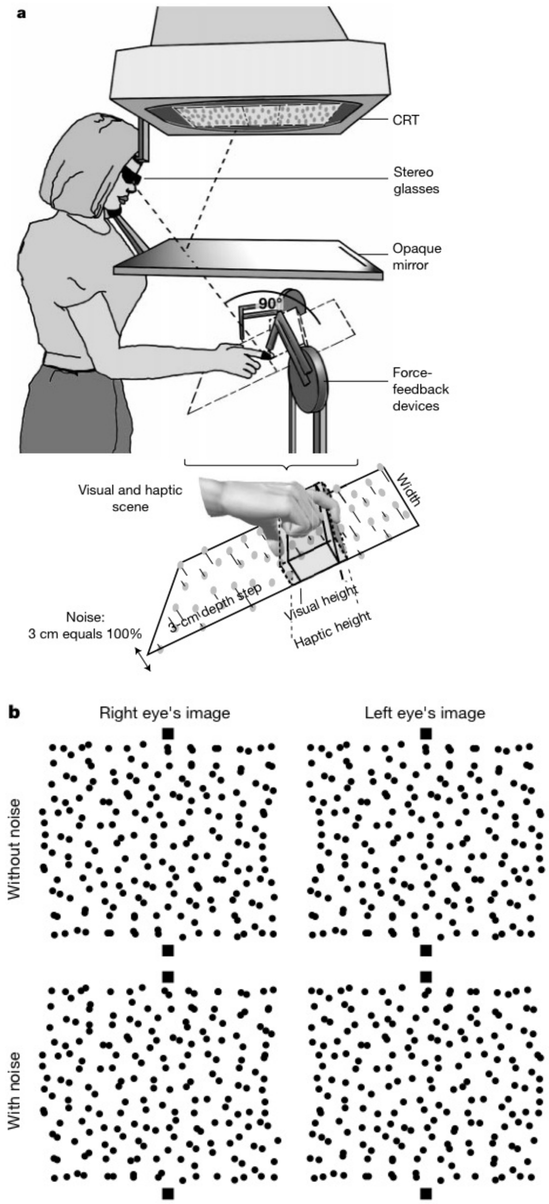
Out[101]= $\mu\to\frac{r_1\,x_1}{r_1+r_2}+\frac{r_2\,x_2}{r_1+r_2}$

It follows that the combined estimate of the averages is the weighted sum of the separate estimates, where the weights w_i are determined by the relative reliabilities :

$$\hat{\mu}_{\text{combined}}=\hat{\mu}_{\text{cue1}}\,w_1+\hat{\mu}_{\text{cue2}}\,w_2=\hat{\mu}_{\text{cue1}}\,\frac{r_1}{r_1+r_2}+\hat{\mu}_{\text{cue2}}\,\frac{r_2}{r_1+r_2}.$$

An application to integrating cues from vision and haptics (touch)

When a person looks and feels an object the two cues typically combine to form one perceived size. Vision often dominates the integrated percept--e.g. the perceived size of an object is driven more strongly by vision than by touch. Why is this? Ernst and Banks showed that the reliability of the visual and haptic information determines which cue dominates. They first measured the variances associated with visual and haptic estimation of object size. They used these measurements to construct a maximum-likelihood estimator that integrates both cues. They concluded that the nervous system combines visual and haptic information in a fashion that is similar to a maximum-likelihood ideal observer. Specifically, visual dominance occurs when the variance associated with visual estimation is lower than that associated with haptic estimate.



See Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.

Perceptual explaining away, cooperative computation

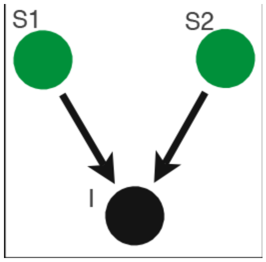
We will now describe more complex examples of integration in which vision

Perception as puzzle solving

Rock, I. (1983). *The Logic of Perception*. Cambridge, Massachusetts: M.I.T. Press.

Perceptual explaining away

Both causes S1 and S2 can be primary variables.



The above graph describes the factorization:

$$p(S1,S2,I) = p(I|S2,S2)\ p(S1)p(S2)$$

If we average over I, S1 and S2 are independent. However, knowledge of the value of I makes S1 and S2 conditionally dependent. The two causes S1 and S2 can behave like competing hypotheses to explain the data I.

In general, “explaining away” is a phenomenon that occurs in probabilistic belief networks in which two (or more) variables influence a third variable whose value can be measured (Pearl, 1988). Once measured, it provides evidence to infer the values of the influencing variables.

Imagine two fair coins that can be flipped independently, and the results (heads or tails) have an influence on a third variable. For concreteness, assume the third variable’s value is 1 if both coins agree, and 0 if not (a logical NOT-XOR function). If we are ignorant of the value of the third variable, knowledge of one influencing variable doesn't help to guess the value of the other—the two coin variables are independent. (This is called marginal independence, “marginal” with respect to the third variable, I)

But if the value of the third variable is measured (e.g. suppose we look and see it is 1), the two coin variables become coupled-- they are *conditionally dependent*. Now knowing that one coin is heads guarantees that the other one is too. Although we still can’t perfectly predict the values of the coins, we now know something about them we didn’t know before.

Now imagine a slight twist on the problem. Suppose you are most interested in the value of one of the coin flips (C1), not the other (C2). If you have any additional (auxiliary) evidence that the other coin’s value (C2) is say, probably “heads”, then an optimal guess would be to say C1 is heads too.

The phrase “explaining away” arises because coupling of variables through shared evidence often arises in human reasoning, when the influences can be viewed as competing causes. A change in belief of one of the competing hypotheses changes the belief in the other. Human reasoning is particularly good at these kinds of inferences.

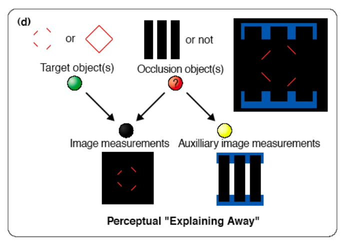
“Explaining away” is also a characteristic of perceptual inferences, for example when there are alternative perceptual groupings consistent with a set of identical or similar sets of local image features.

Demonstrations of explaining away in perception

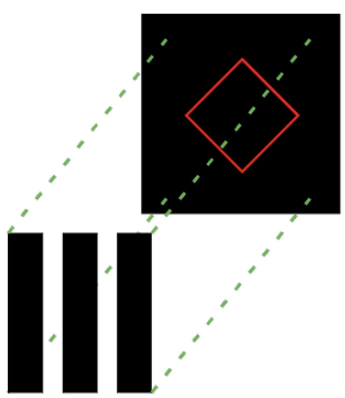
Several perceptual phenomena that we've seen before can be interpreted as "explaining away".

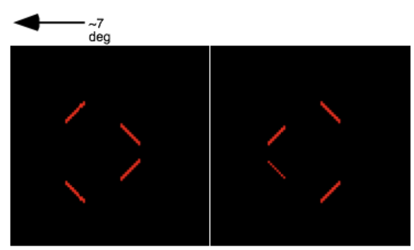
Translating diamond with "occluding occluders"

A strong argument for a process that does “explaining away” is human vision’s adeptness at solving occlusion problems.



Occlusion & motion. Recall the translating diamond used to illustrate the aperture problem. Perception doesn’t even need strong occlusion cues to arrive at the conclusion “diamond with missing vertices”. When the diamond is seen as coherently translating, one often also interprets the vertices as being covered by rectangular occluders (Lorenceau & Shiffrar).

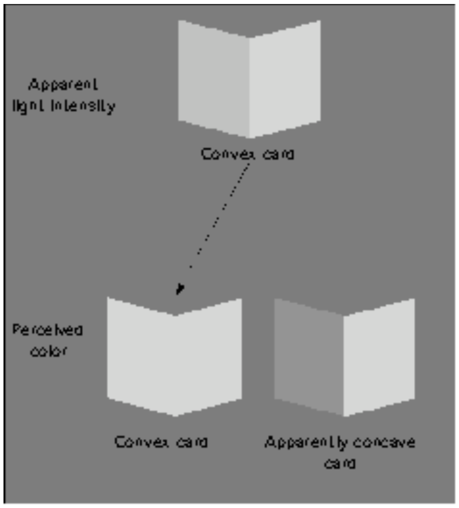




▸ 1. Can you make a stereo analog of the translating diamond with occluders?

Lightness & surface geometry

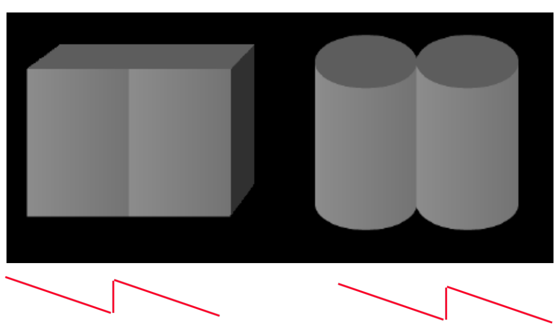
19th century demo: Mach card



Lightness and shape

Recall the lightness demonstration that is similar to the Craik-O'Brien-Cornsweet effect, but difficult to explain with a simple filter mechanism (Knill, D. C., & Kersten, D. J., 1991). The idea is that the lightness of a pair of luminance gradients on the left of the figure below look different, whereas they look similar for the pair luminance gradients on the right. The reason seems to be due to the fact that the luminance gradients on the right are attributed to smooth changes in shape, rather than smooth changes in illumination. http://vision.psych.umn.edu/www/kersten-lab/demos/lightness.html

These demonstrations suggest the existence of scene representations in our brains for shape, reflectance and light source direction.



► 2. Draw a diagram to illustrate the above illusion in terms of "explaining away"

Dependence of lightness on spatial layout

Gilchrist:

In the 1970's, Alan Gilchrist was able to show that the lightness of a surface patch may be judged either dark-gray, or near-white with only changes in perceived spatial layout (Gilchrist, A. L. (1977). How did he do this? What is going on? Interpret lightness as reflectance estimation.

The figures below illustrate Gilchrist's Room-in-a-Shoe-Box experiment and the Coplanar card experiment.

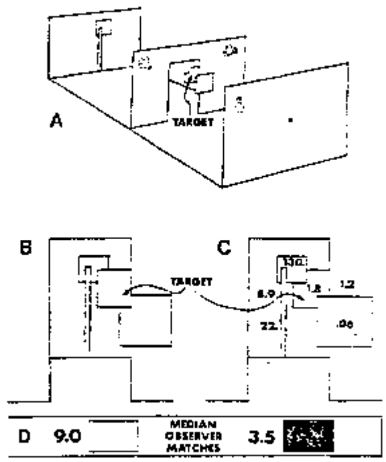


Figure 5. (A) Perspective view of the parallel planes display, showing hidden light bulbs. The display (as seen through the pinhole) in which the target appeared to be located either (B) in the near plane or (C) in the far plane, with luminances shown in foot-Lamberts. (D) The average match from a Munsell chart for the two displays.

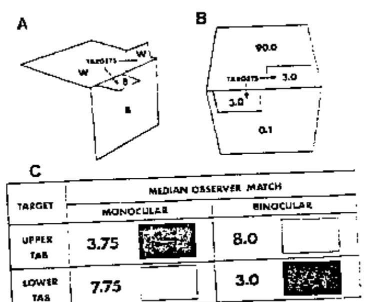


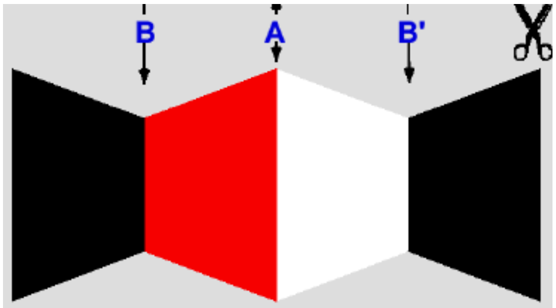
Figure 4. (A) Perspective view of the stimulus display used in the critical test, showing color (B, black; W, white) of each part. (B) Monocular retinal pattern showing luminances in foot-Lamberts. (C) Average Munsell matches for monocular and binocular viewing conditions.

Color & shape

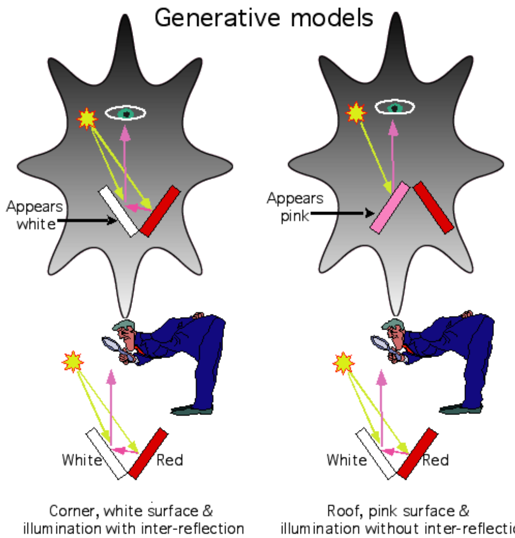
Recall the color card experiment (Bloj, Kersten & Hurlbert)

Demo

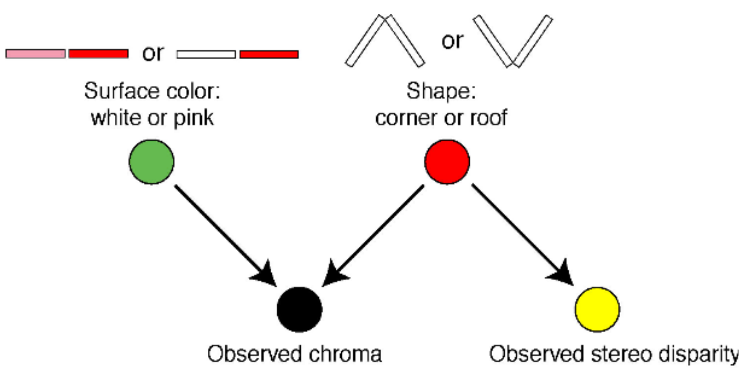
http://gandalf.psych.umn.edu/users/kersten/kersten-lab/Mutual_illumination/BlojKerstenHurlbertDemo99.pdf



Interpretation



Interreflection as explaining away. Stereo can be used as an auxiliary cue to change the perceived shape from concave to convex.



Dependence of shape on perceived light source direction

Dependence of shape on perceived light source direction

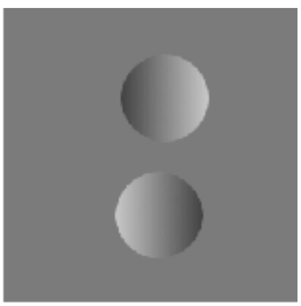
Brewster (1926), Gibson, Ramachandran, V. S. (1990), crater illusion and the single light source assumption

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. Nat Neurosci, 7(10), 1057-1058.

Sufficient evidence can overcome the prior. See: Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. PNAS, 108(30), 12551–12553.

Vertical light direction

Horizontal light direction



Transparency

Structure from motion and transparency

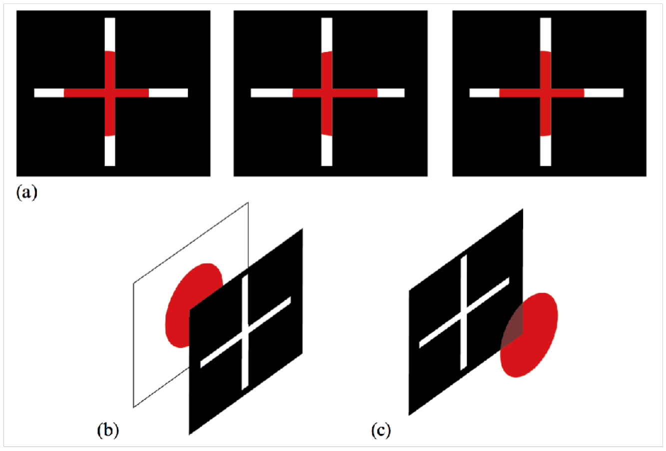
Dependence of transparency on perceived depth

Recall the transparency and depth from motion demonstrations.

(See Kersten et al., 1992) http://gandalf.psych.umn.edu/users/kersten/kersten-lab/demos/transparency.html

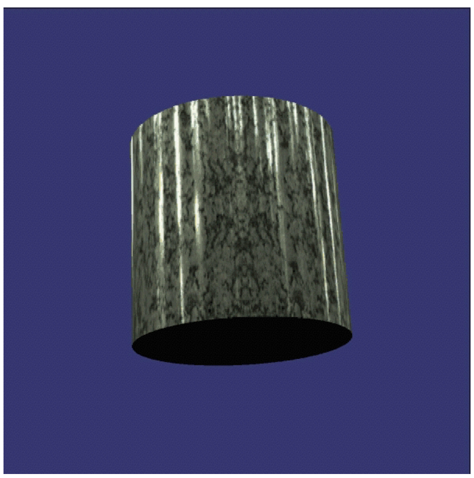
Transparency and depth from stereo demos, neon color spreading

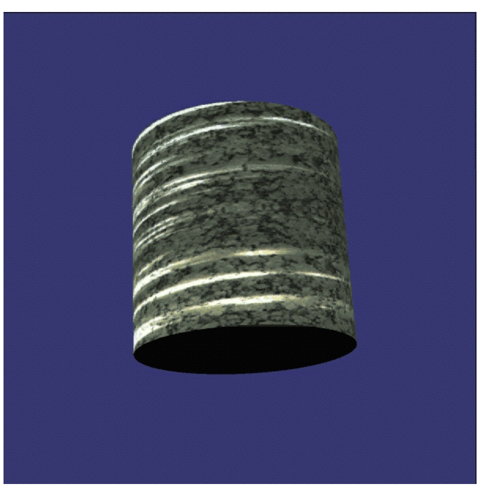
Nakayama, Shimojo (1992); Nakayama, K., Shimojo, S., Anderson, B. L., & Kramer, P. (2009). Nakayama, Shimojo, and Ramachandran's 1990 paper. Perception-London. http://doi.org/10.1068/ldmk-nak



Perception of material gloss depends on curvature

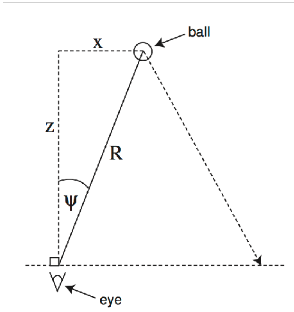
(Bruce Hartung & Dan Kersten)



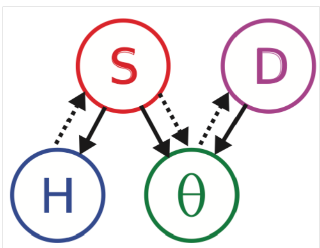


Feeling the size of an object can improve subsequent visual trajectory estimation

In order to intercept a ball at the right location, the visual system has to decide if it is looking at a small object that is near, or a large object that is far.

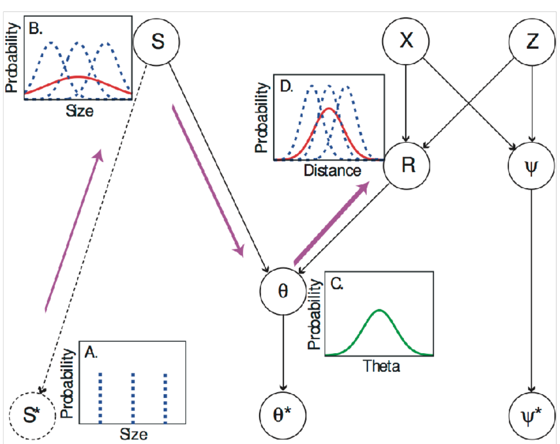


Battaglia, P. W., Schrater, P. R., & Kersten, D. J. (2005) showed that humans could incorporate haptic sensing of an objects 3D size to improve interception.
This suggested that visual motor estimations could discount variations in 3D object size contributions to image size. And that haptic information (H) about physical size S, could explain away these variations in image size θ that are caused by both S and depth D.



The above figure is from: Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. PLoS Computational Biology, 7(6), e1002080. doi:10.1371/journal.pcbi.1002080

The following figure fleshes out more details showing the variable, ψ, that most directly influences participants interception accuracy (Battaglia, Schrater, & Kersten, 2005).



In this figure S^*, θ^*, ψ^* represent measurements of physical size, angular size, and angle required for interception along the x-axis (dotted line in first figure). R represents distance.

Example from computer vision: image parsing

Incorporating higher-level knowledge--Image parsing and recognition using cooperative computation

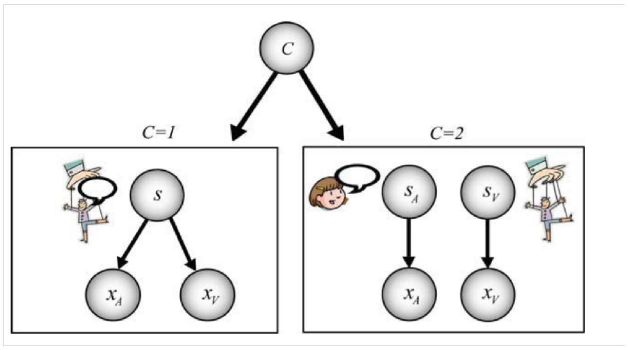
In limited domains, feedforward computations can solve object recognition in natural images, albeit with errors. One way to reduce the errors is to have a feedback pass that tries to “predict” the input using a model of synthesis. In the example below, the algorithm “knows” about text and faces. Everything else is “clutter”. But it “knows” clutter too--it assumes clutter is a generica texture with pre-determined statistical structure. The first pass detects and recognizes letter characters and faces. It decides there is a face n the tree (rightmost part of first figure below). Based on decisions in the first pass, and its built-in knowledge of the nature of faces, characters and clutter texture, the feedback pass synthesizes what the image “should be”. This second pass “explains away” the face false positive in the tree, as texture.

For explaining away applied to computer vision solutions to segmentation and recognition, see: Tu Z, Zhu S-C (2002), Zhu and Tu (2000). For a review, see: Yuille and Kersten (2006).

Mixture models

Recall the problem of integrating motion information given multiple apertures. We identified two parts to the problem: selection and integration, but postponed the question of selection--which information should be integrated and which should not. Selection processes can be automatic, or task and attention driven.

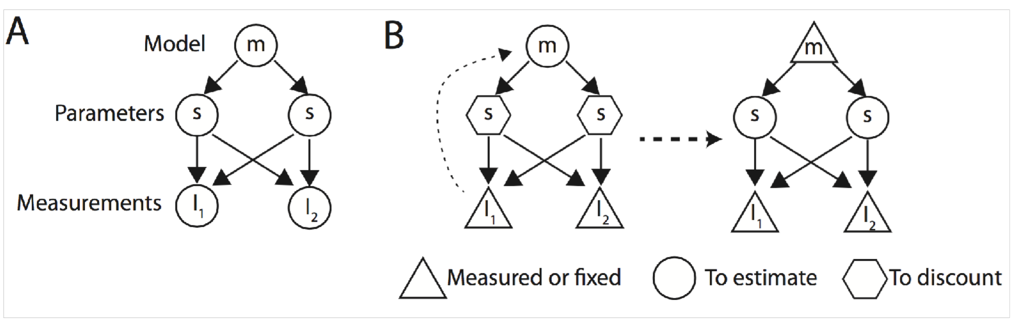
Example from: Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. PLoS ONE, 2(9), e943. http://doi.org/10.1371/journal.pone.0000943.t001



Graphical model interpretations

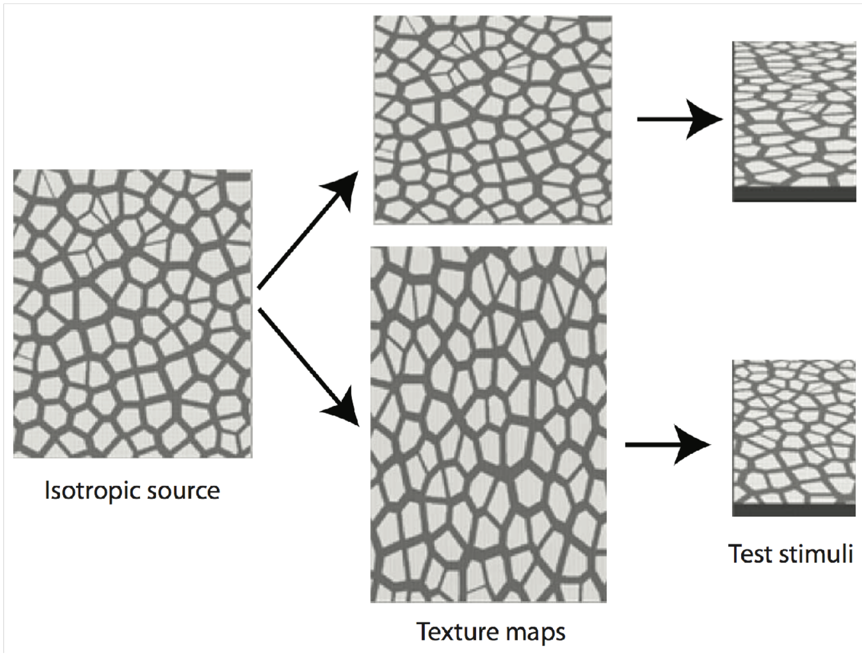
What to integrate out? And in what order?

There are several ways to do inference on a graph several model choices, each with various parameters. For example, given two possible models or “causal explanations”, one could combine the two inferences to do “model averaging”. Or the process could first select the model that is most probable, and then given that model infer the parameters from the data.



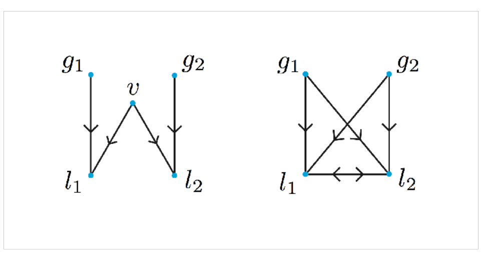
See: Stocker, A. A., & Simoncelli, E. (2008). A Bayesian model of conditioned perception. Advances in Neural Information Processing Systems, 20, 1409–1416.

The first application of mixture models to perceptual behavior was by: Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. Vision Research, 43(7), 831–854. http://doi.org/10.1016/S0042-6989(03)00003-8. In Knill’s example the data, i.e. image texture, could be generated by an isotropic homogeneous texture process, or by an homogeneous texture process only. Knill’s finding was that human vision is biased to interpret image texture as isotropic but if enough data is available the system turns off the isotropy assumption and interprets texture using the homogeneity assumption only.

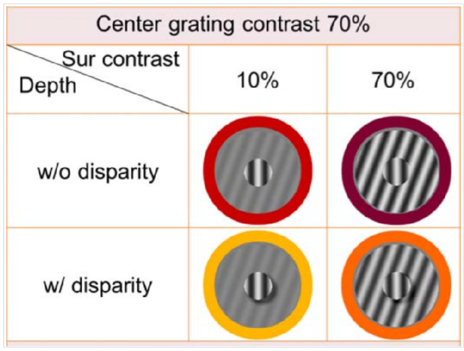


Figure/ground and divisive normalization

The problem of model selection arises very early in the visual system. For example, divisive normalization may be modulated by factors that determine whether spatial filter outputs are coming from common causes in the world or not. For example, given two filter outputs, do they both come from the figure (or surround), or does one come from the figure and the other from the surround? Divisive normalization has been explained as an efficient inference solution to the generative model in the graph on the left. If one integrates out *v*, all the variables become dependent, as shown in the figure on the right.



Qiu, C., Kersten, D., & Olman, C. A. (2013) showed how segmentation cues such as contrast-contrast (between center and surround) and stereo disparity could affect the well-known tilt-illusion. Their results could be explained in terms of decoupling the causes between center and surround(see Schwartz, O., Sejnowski, T. J., & Dayan, P. , 2009). If decoupled, the variable *v* in the above has no common influence on the image measurements l.



See too: Schwartz, O., & Coen-Cagli, R. (2013) and Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015).

References

Cue integration, cooperative computation

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. Nat Neurosci, 7(10), 1057-1058.
Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering Intrinsic Scene Characteristics from Images. In A. R. Hanson, & E. M. Riseman (Ed.), Computer Vision Systems (pp. 3-26). New York: Academic Press.
Battaglia, P. W., Schrater, P. R., & Kersten, D. J. (2005). Auxiliary object knowledge influences visually-guided interception behavior. Proceedings of the 2nd symposium on Applied perception in graphics and visualization, 145–152.
Battaglia, P. (2010). Bayesian perceptual inference in linear Gaussian models. MIT-CSAIL-TR-2010-046.
Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. PLoS Computational Biology, 7(6), e1002080. doi:10.1371/journal.pcbi.1002080
Bergstrom, S. S. (1977). Common and Relative Components of Reflected Light as Information About the Illumination, Colour and Three-Dimensional Form of Objects. 18, 180-186).
Bishop, C. M. (2006). Pattern recognition and machine learning. New York: Springer.
Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. Edinburgh Journal of Science, 4, 99-108.
Clark, J. J., & Yuille, A. L. (1990). Data Fusion for Sensory Information Processing . Boston: Kluwer Academic Publishers.
Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. Nature Neuroscience, 18(11), 1648–1655. http://doi.org/10.1038/nn.4128
Ernst MO, Banks MS, Bulthoff HH (2000) Touch can change visual slant perception. Nat Neurosci 3:69-73.
Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.
Gilchrist, A. L. (1977). Perceived Lightness Depends on Perceived Spatial Arrangement. Science, 195, 185-187.
Gibson, J. J. (1950). The Perception of the Visual World. Boston, MA: Houghton Mifflin.
Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion

within, but not between, senses. Science 298:1627-1630.
Humphrey, K. G., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., & Menon, R. S. (1996). Differences in Perceived Shape from Shading Correlate with Activity in Early Visual Areas.,, 1-16.
Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3, 79-87.
Jacobs RA (2002) What determines visual cue reliability? Trends Cogn Sci 6:345-350.
Kersten, D. J. (1991). Transparency and the Cooperative Computation of Scene Attributes. In M. Landy, & A. Movshon (Ed.), Computational Models of Visual Processing. Cambridge, Massachusetts: M.I.T. Press.
Kersten, D., Bülthoff, H. H., Schwartz, B., & Kurtz, K. (1992). Interaction between transparency and structure from motion. Neural Computation, 4(4), 573-589.
Kersten D, Yuille A (2003) Bayesian models of object perception. Current Opinion in Neurobiology 13:1-9.
Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian Inference. Annual Review of Psychology 55:271-304.
Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. Nature, 351, 228-230.
Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. Vision Research, 43(7), 831–854.
Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. PLoS ONE, 2(9), e943. http://doi.org/10.1371/journal.pone.0000943.t001
Mach, E. (1886, 1959). The Analysis of Sensations . New York: Dover.
Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. Science, 257, 1357-1363.
Pearl J (1988) Probabilistic reasoning in intelligent systems : networks of plausible inference, Rev. 2nd printing. Edition. San Mateo, Calif.: Morgan Kaufmann Publishers.
Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. Nature, 317, 314-319.
Poggio, T., Gamble, E. B., & Little, J. J. (1988). Parallel integration of vision modules. Science, 242, 436-440.
Qiu, C., Kersten, D., & Olman, C. A. (2013). Segmentation decreases the magnitude of the tilt illusion. Journal of Vision, 13(13), 19–19. http://doi.org/10.1167/13.13.19
Ramachandran, V. S. (1990). Visual perception in people and machines. In A. Blake, & T. Troscianko (Ed.), A.I. and the Eye John Wiley & Sons Ltd.
Ripley, B. D. (1996). Pattern Recognition and Neural Networks . Cambridge, UK: Cambridge University Press.
Schrater PR, Kersten D (2000) How optimal depth cue integration depends on the task. International Journal of Computer Vision 40:73-91.
Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. Journal of Vision, 9(4), 19.1–20. http://doi.org/10.1167/9.4.19.
Schwartz, O., & Coen-Cagli, R. (2013). Visual attention and flexible normalization pools. Journal of

Vision, 13(1), 25–25. http://doi.org/10.1167/13.1.25
Stocker, A. A., & Simoncelli, E. (2008). A Bayesian model of conditioned perception. Advances in Neural Information Processing Systems, 20, 1409–1416.
Tjan B., Braje, W., Legge, G.E. & Kersten, D. (1995) Human efficiency for recognizing 3-D objects in luminance noise.Vision Research, 35, 3053-3069.
Todd, J. T., & Mingolla, E. (1983). Perception of Surface Curvature and Direction of Illumination from Patterns of Shading. Journal of Experimental Psychology: Human Perception & Performance, 9(4), 583-595.
Tu Z, Zhu S-C (2002) Parsing Images into Region and Curve Processes. In: Proc. of the 7th European Conference on Computer Vision, p 393 ff. Copenhagen, Denmark: Springer-Verlag, Berlin Heidelberg.
von der Heydt R, Friedman H, Zhou HS (2003) Searching for the neural mechanisms of color filling-in. In: Filling-in: From Perceptual Completion to Cortical Reorganization (Pessoa L, P DW, eds), pp 106-127. Oxford: Oxford University Press.
Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In K. D.C., & R. W. (Ed.), Perception as Bayesian Inference Cambridge, U.K.: Cambridge University Press.
Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci, 10(7), 301-308. http://gandalf.psych.umn.edu/~kersten/kersten-lab/papers/yuillekerstenTICs2006.pdf
Zhu S-C, Zhang R, Tu Z (2000) Integrating Bottom-up/Top-Down for Object Recognition by Data Driven Markov Chain Monte Carlo. In: Proc. of Int'l Conf. on Computer Vision and Pattern Recognition. SC.