

# Cue Coupling

A.L. Yuille and D. Kersten

## Abstract

### I. VISION MODULES AND CUE COMBINATION

To what extent is the visual system built by combining modular components? This issue was discussed in the basic architecture section and viewpoints differed between advocates of separate pathways encoding different feature dimensions [9], [17] and other who argued against separation because most cells are sensitive to all dimensions [16] and, premature separation, raises concerns about how different pathways can be combined in order to yield a unified percept. At the functional/behavioral level, psychophysicists have studied how humans combine different visual cues – such as shading, texture, binocular stereo, structure from motion – to get depth. Marr [20] invoked the principle of modular design and proposed that different cues could be processed by semi-independent modules and then combined into representations – e.g., Marr’s 2 1/2 D sketch combined different cues for depth into a single representation. Computer vision researchers have also tended to study these cues in isolation and developed algorithms from them individually.

We argue that although there is evidence that cues can act independently (and deficit studies show that damage to localized brain regions is able to knock out some visual cues while leaving others unimpaired). Nevertheless, evidence suggests that the modules must be capable of tightly interacting in certain situations. In particular, there is strong evidence that high-level recognition affects the estimation of three-dimensional shape (e.g., a rigidly rotating inverted face mask is perceived as non-rigidly deforming face, while most rigidly rotating objects are perceived to be rigid). It is also clear that, despite the partial success of the models described for segmentation in section (??) (and their more sophisticated descendants), the ability to perform segmentation relies partially on context and the ability to do object detection, hence involving some top-down processing.

This section gives an overview of visual cues and the strategies for combining them. Clark and Yuille [4] argued that cue coupling should be formulated in terms of Bayesian probability theory so that the uncertainties of the cues could be taken into account and their statistical dependencies made explicit (previous authors, e.g., Marr [20], had not specified the details of how cues should be combined). Clark and Yuille [4], see also [24], divided cues coupling into two types: (i) ‘weak coupling’ which corresponded to combining independent cues and which often reduced to combining cues by weighted averaging, and (ii) ‘strong coupling’ which required more sophisticated methods such as model selection, or ‘competitive priors’, to model phenomena where small changes of one cue can dramatically change the percept. Since then many studies have shown that humans often couple cues weakly in an optimal sense – i.e., that they weight the cues based on their reliability.

To understand this it helps to consider how the image is formed from the structure of the real world. We can think of this in terms of causes. Different factors in the scene combine to cause the image. These combinations can be complicated and are rarely independent – as they would need to be for weak coupling to be appropriate. In addition, some visual cues are only valid under special conditions. For example, classic shape from texture methods assume that there are texture elements of roughly similar sizes on a surface in the scene. This assumption can be relaxed but nevertheless there are many places in images where shape from texture cues simply do not exist. Similarly, classic shape from shading models apply only to Lambertian shading models with known albedo and a single light source.

It is helpful to think of cue combination in terms of graphical models as we illustrate in figure (6). We stress that these diagrams are conceptual and ignore the details of the models – i.e. each node should

correspond to a lattice of nodes. Figure (6)(B) shows three conditions: (i) the simplest case when a single cue is present, (ii) where two factors combine to cause an image, and (iii) where there is *common cause* of two cues. We show two examples of the later two cases in figure (1).

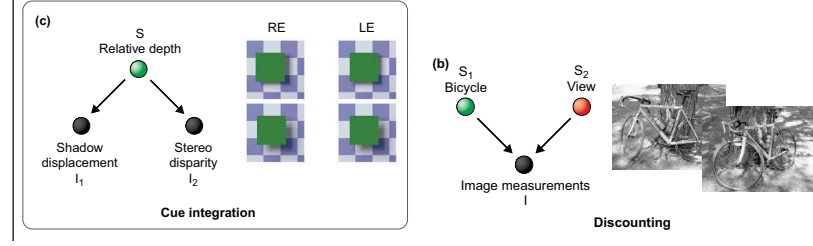


Fig. 1. Left Panel: An example of common cause. The shading and binocular stereo cues are caused by the same event – two surface with one partially occluding the other. Right Panel: The image of the bicycle is caused by the pose of the bicycle, the viewpoint of the camera, and the lighting conditions.

If it is important to estimate one of the (parent) state variables accurately, the other one needs to be integrated out, i.e. discounted. Invariance is the flip side of discounting. If both state variables are important, they can behave as competing hypotheses either of which can explain the data, called “explaining away”. The parent values are said to be marginally independent (summing over  $I$ ), but conditionally dependent once given a measurement. Figure 6B (right) illustrates the case when one cause leads to two effects. The graph implies that the two measurements are conditionally independent given  $s$ . This generative model is the basis for tests of optimal cue integration. A distinctive prediction of Bayesian observers is that decisions should be based on full knowledge of the posterior distribution.

Figure 6D & E illustrate a richer generative structure, where there is more than one type of model,  $m$ , that could have produced the measurements. Each model may have its own distinct set of parameters. As we will describe in subsection (I-B) Knill showed that, when estimating surface orientation from texture cues, the visual system uses different models depending on evidence in the image [12]. The visual system usually interprets a 2D texture as caused by an underlying isotropic 3D texture. However, textures may also be anisotropic. Human surface judgments were well-modeled by a Bayesian observer that assumes that surfaces in the world can come in the two types.

For the generative structure in Figure 6D, task choices—what to integrate out—becomes a critical modeling question. What if there is more than one model for the brain to consider? Formally, which model can be decided by integrating out the proximal causes  $s$ , to specify the posterior,  $p(m|I)$ , on more distal causes or “models”  $m$  that represents the discrete set of model choices. With a generative structure such as shown in Figure 6D, one could estimate the most probable model AND parameters consistent with the measurements. Alternatively, one could integrate out the model variables, and find the most probable parameters. Another study showed that human localization based on combining sound and light behaves as if it can infer the causal structure (i.e. which kind of graph best explains the data, see Figure 6E), and then integrate or not integrate cues accordingly [14]. It has also been proposed that humans may choose a causal structure in proportion to its probability, a strategy in the cognitive decision making literature known as probability matching [7]. As described in subsection (I-B), Shams et al. [22] provide evidence that human observers localization judgments were most consistent with probability matching.

### A. Weak Coupling

Weak coupling methods assume that the visual cues for depth are modular. This means that they function independently and then combine their outputs. This combination takes into account the uncertainty in the cues. If the cues are modeled using Gaussian distributions then this leads to linear weighted averaging.

Early testing of weak coupling models were qualitative but limited [2], [15],[25]. But there is now considerable evidence for weak coupling. Jacobs [10], and Ernst & Banks [5], there have been numerous

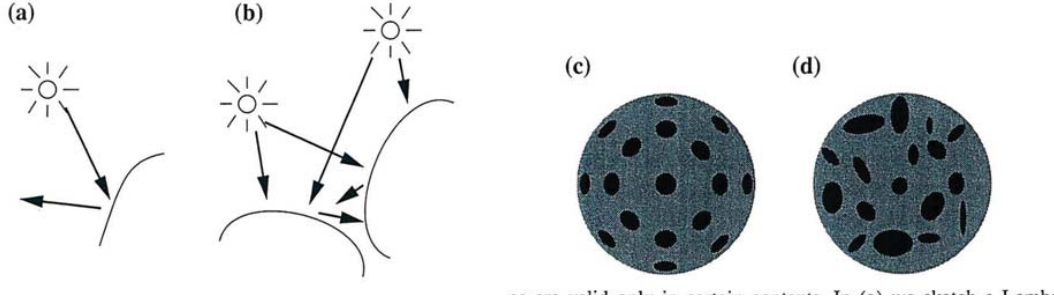


Fig. 2. Model selection may need to be applied in order to decide is a cue can be used. Shape from shading cues will work for case (a) because the shading pattern is simple == a smooth convex surface illuminated by a single source. But for case (b) the shading pattern is complex – due to mutual reflection between the two surfaces – and so shape from shading cues will be almost impossible to use. Similarly, shape from texture is possible for case (c) because the surface contains a regular texture pattern but is much harder for case (d) because the texture is irregular.

studies that have tested whether humans combine sensory information weighted by reliability. The majority of these studies confirm optimality, with interesting exceptions [3], [6]. Most studies of cue integration have been restricted to continuous-valued measurements whose precision is modeled using standard gaussian and independence assumptions, and thus a linear weighting function.

To understand how cue coupling can yield linear weighted summation consider a simple model where the cues  $\vec{C}_1, \vec{C}_2$  were generated by independent distribution  $P(\vec{C}_1|\vec{S})P(\vec{C}_2|\vec{S})$  where both distributions are Gaussians.

$$P(\vec{C}_1|\vec{S}) = \frac{1}{Z_1} \exp\left\{-\frac{|\vec{C}_1 - \vec{S}|^2}{2\sigma_1^2}\right\}, \quad P(\vec{C}_2|\vec{S}) = \frac{1}{Z_2} \exp\left\{-\frac{|\vec{C}_2 - \vec{S}|^2}{2\sigma_2^2}\right\}. \quad (1)$$

The optimal estimation  $\vec{S}^* = \arg \max P(\vec{C}_1|\vec{S})P(\vec{C}_2|\vec{S})$  is given by:

$$\vec{S}^* = \frac{1}{1 + (\sigma_1^2/\sigma_2^2)} \vec{C}_1 + \frac{1}{1 + (\sigma_2^2/\sigma_1^2)} \vec{C}_2. \quad (2)$$

This shows that the optimal combined is a weighted linear sum of the two cues. The cue with the smallest variance will be weighted more highly (i.e. cue  $\vec{C}_1$  if  $\sigma_1^2 < \sigma_2^2$ ).

### B. Strong Coupling

Weak coupling is based on a modularity assumption which is often not valid (of course, the visual system may assume modularity even if it is sub-optimal). There are a range of classic experiments which are inconsistent with weak coupling [2], [1], [11]. Yuille and Bulthoff [25] interpreted these as variants of strong coupling using the Bayesian framework developed in [4].

For example, there may be alternative causes which may compete to explain the image and small changes in the image can lead to very different interpretations. The work by Blake and Bulthoff [1] is a classic example where a sphere has a Lambertian (diffuse) reflection function and is viewed binocularly, see figure (2)(left). A specular component is adjusted so that it can lie in front of, between the center and the sphere, or at the center of the sphere. If the specularity lies at the center then it is perceived as a light bulb and the sphere is perceived to be transparent. If the specularity is placed at the right position between the center and the sphere, then the sphere is perceived to be glossy. If the specularity lies in front of the sphere then it is seen as a cloud floating in front of a matte (diffuse, Lambertian sphere). In other experiments [11] there are two surfaces which can be seen to either move rigidly together or to move independently. Either percept can be obtained by making small variations to the transparency cues, see figure (3)(right). Yuille and Bulthoff [25] interpreted both these phenomena as strong coupling with competitive priors.

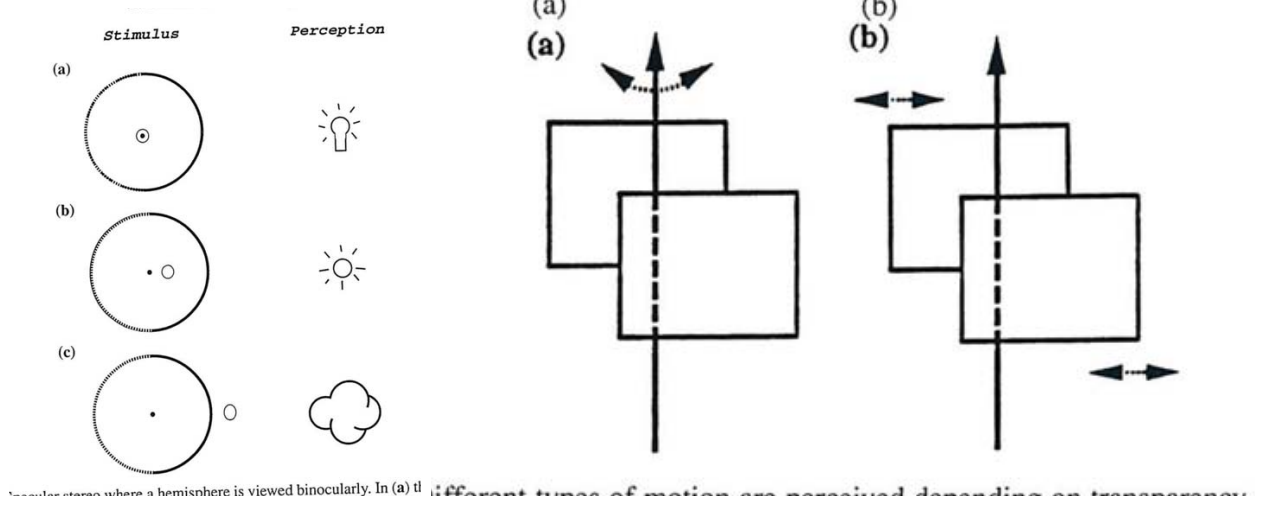


Fig. 3. Examples of strong coupling with competitive priors. A sphere is viewed binocularly (left) and small changes in the position of the specularities lead to very different percepts (Blake and Bülthoff 1990). Similarly altering the transparency of the moving surfaces (right) can make the two surfaces appear to rotate either rigidly together or independently.

### Examples of Strong Coupling

This section gives two examples of strong coupling. The first example concerns the perception of texture while the second example deals with coupling different modalities.

The first example is by Knill and concerns the estimating of depth from texture cues [12]. This relates to competitive priors because there are several alternative models for generating the image and the human observer must infer which is most likely. More formally, the data is generated by a mixture of models which enables non-linear cooperative interaction interactions between cues. In this example the data could be generated by isotropic homogeneous texture or by homogeneous texture only. Knill's findings is 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.

The posterior probability distribution for  $S$  is given by:

$$P(S|I) = \frac{P(I|S)P(S)}{P(I)}, \quad P(I|S) = \sum_{i=1}^n \phi_i P_i(I|S), \quad (3)$$

where  $\phi_i$  is prior probability of model  $i$ , and  $p_i(I|S)$  is corresponding likelihood function.

More specifically, texture features  $T$  can be generated by either an isotropic surface or a homogeneous surface. The surface is parameterized by tilt and slant  $\sigma, \tau$ . Homogenous texture is described by two parameters  $\alpha, \theta$  and isotropic texture is a special case where  $\alpha = 1$ . This gives two likelihood models for generating the data:

$$P_h(T|(\sigma, \tau), \alpha, \theta), \quad P_i(T|(\sigma, \tau), \theta) \quad (4)$$

Here  $P_i(T|(\sigma, \tau), \theta) = P_h(T|(\sigma, \tau), \alpha = 1, \theta)$ .

Isotropic textures are a special case of homogenous textures (also rigid motion is a special class of non-rigid motion). The homogeneous model has more free parameters and hence has more flexibility to fit the data which suggests that human observers should always prefer it. But the Occam factor [19] means that this advantage will disappear if we put priors  $P(\alpha)P(\theta)$  on the model parameters and integrate them out. This gives:

$$P_h(T|(\sigma, \tau)) = \int \int d\alpha d\theta P_h(T|(\sigma, \tau), \alpha, \theta), \quad P_i(T|(\sigma, \tau)) = \int d\theta P_h(T|(\sigma, \tau), \theta) \quad (5)$$

Integrating over the model priors smooths out the models. The more flexible model,  $P_h$ , has only a fixed amount of probability to cover a large range of data (e.g. all homogeneous textures) and hence has lower probability for any specific data (e.g. isotropic textures).

Knill describes how to combine these models using model averaging. The combined likelihood function is obtained by taking a weighted average:

$$P(T|(\sigma, \tau)) = p_h P_h(T|(\sigma, \tau)) + p_i P_i(T|(\sigma, \tau)), \quad (6)$$

Where  $(p_h, p_i)$  are prior probabilities that the texture is homogeneous or isotropic. We use a prior  $P(\sigma, \tau)$  on the surface and finally achieve a posterior:

$$P(\sigma, \tau|I) = \frac{P(I|(\sigma, \tau))P(\sigma, \tau)}{P(I)}. \quad (7)$$

This model has a rich interpretation. If the data is consistent with an isotropic texture then this model dominates the likelihood and strongly influences the perception. Alternatively, if the data is consistent only with homogeneous texture then this model dominates. This gives a good fit to human performance [12].

The second example involves multisensory integration with structural uncertainty. Human observers are sensitive to both visual and auditory cues. Sometimes these cues have a common cause – e.g., you see a dog moving and hear it barking. In other situations the auditory and visual cues are due to different causes – e.g., a cat moves and a nearby dog barks (we ignore the possibility that the dog’s barking is caused by the cat moving, or vice versa). Ventriloquists are able to fake these interactions by making the audience think that a puppet is speaking by associating the sound (produced by the ventriloquist) with the movement of the puppet. The Ventriloquism effect occurs when visual and auditory cues have different causes – and so are in conflict – but the audience perceive them as having the same cause.

Körding and his collaborators (Körding et al. 2007) developed an ideal observer model which determines whether two cues have a common cause or not. They formulated this using a meta-variable  $C$ , see figure (4). The common cause condition  $C = 1$  means that the positions of the cues  $x_A, x_V$  are generated by the same process  $S$ , see figure (4)(left), by a distribution  $P(x_A, x_V|S) = P(x_A|S)P(x_V|S)$ . Here  $P(x_A|S)$  and  $P(x_V|S)$  are normal distributions  $N(x_A|S, \sigma_A^2), N(x_V|S, \sigma_V^2)$  – with the same mean  $S$  and variances  $\sigma_A^2, \sigma_V^2$ . It is assumed that the visual cues are more precise than the auditory cues so that  $\sigma_A^2 < \sigma_V^2$ . The true position  $S$  is drawn from a probability distribution  $P(S)$  which is assumed to be a normal distribution  $N(0, \sigma_p^2)$ . By contrast,  $C = 2$  means that the cues are generated by two different processes  $S_A$  and  $S_V$ , in which case we have  $P(x_A|S_A)$  and  $P(x_V|S_V)$  which are both Gaussian  $N(S_A, \sigma_A^2)$  and  $N(S_V, \sigma_V^2)$ , see figure (4)(right). We assume that  $S_A$  and  $S_V$  are independent samples from the normal distribution  $N(0, \sigma_p^2)$ . Note that this model involves model selection, between  $C = 1$  and  $C = 2$ , and so in vision terminology is a form of strong coupling and competitive priors (Yuille and Bülthoff 1996).

This model was compared to experiments where brief auditory and visual stimuli were presented simultaneously with varying amount of spatial disparity. Subjects were asked to identify the spatial location of the cue and/or whether they perceive a common cause [23]. The closer the visual stimulus was to the audio stimulus the more likely subjects perceived a common cause. In this case subjects’ estimate of its position is strongly biased by the visual stimulus (because it is considered more precise with  $\sigma_V^2 > \sigma_A^2$ ). But if subjects perceive distinct causes then their estimate is pushed away from the visual stimulus and exhibits *negative bias*. Körding *et al.* (Körding et al. 2007) argue that this bias is a selection bias stemming from restricting to trials in which causes are perceived as being distinct. For example, if the auditory stimulus is at the center and the visual stimulus at 5 degrees to right of center – then sometimes the (very noisy) auditory cue will be close to the visual cue and hence judged to have a common cause while on other cases the auditory cause will be further away (more than 5 degrees). Hence the auditory cue will have a truncated Gaussian (if judged to be distinct) and will yield negative bias.

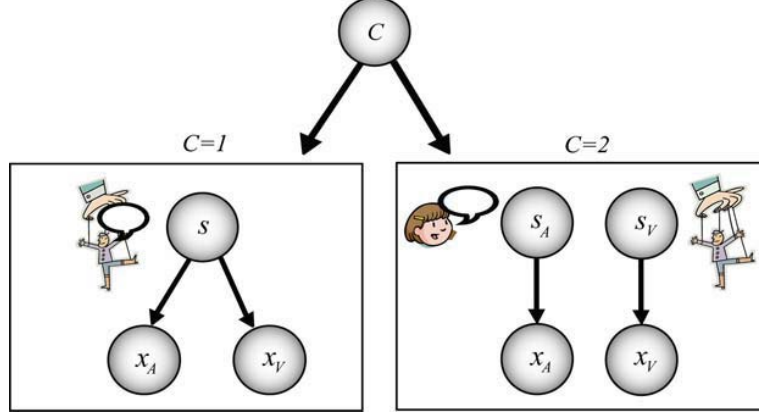


Fig. 4. The subject is asked to estimate the position of the cues and to judge whether the cues are from a common cause – i.e. at the same location – or not. In Bayesian terms the task of judging whether the cause is common can be formulated as model selection – are the auditory and visual cues more likely to be generated from a single cause (left) or by two independent causes (right).

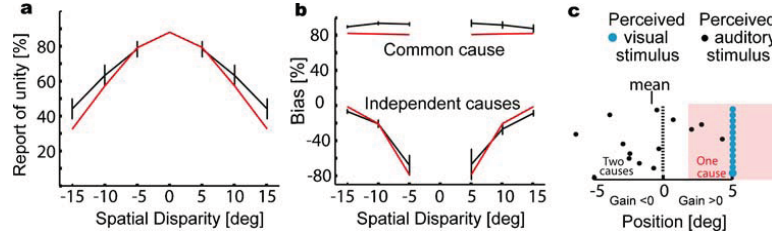


Fig. 5. Reports of causal inference. a) The relative frequency of subjects reporting one cause (black) is shown (reprinted with permission from [15]) with the prediction of the causal inference model (red). b) The bias, i.e. the influence of vision on the perceived auditory position is shown (gray and black). The predictions of the model are shown in red. c) A schematic illustration explaining the finding of negative biases. Blue and black dots represent the perceived visual and auditory stimuli, respectively. In the pink area people perceive a common cause.

More formally, the beliefs  $P(C|x_A, x_V)$  in these two hypotheses  $C = 1, 2$  are obtained by summing out the estimated positions  $s_A, s_B$  of the two cues as follows:

$$\begin{aligned}
 P(C|x_A, x_V) &= \frac{P(x_A, x_V|C)P(C)}{P(x_A, x_V)} \\
 &= \frac{\int dS P(x_A|S)P(x_V|S)P(S)}{P(x_A, x_V)}, \quad \text{if } C = 1, \\
 &= \frac{\int \int dS_A dS_V P(x_A|S_A)P(x_V|S_V)P(S_A)P(S_V)}{P(x_A, x_V)}, \quad \text{if } C = 2.
 \end{aligned} \tag{8}$$

There are two ways to combine the cues. The first is *causal selection*. This estimates the most probable model  $C^* = \arg \max P(C|x_V, x_A)$  from the input  $x_A, x_V$  and then uses this model to estimate the most likely positions  $s_A, s_V$  of the cues from the posterior distribution:

$$P(s_V, s_A) \approx P(s_V, s_A|x_V, x_A, C^*) = \frac{P(x_V, x_A|s_V, s_A, C^*)P(s_V, s_A|C^*)}{P(x_V, x_A|C^*)} \tag{9}$$

The second way to combine the cues is by *causal averaging*. This does not commit itself to choosing  $C^*$  but instead averages over both models:

$$P(s_V, s_A | x_V, x_A) = \sum_C P(s_V, s_A | x_V, x_A, C) P(C | x_V, x_A) = \sum_C \frac{P(x_V, x_A | s_V, s_A, C) P(s_V, s_A | C) P(C | x_V, x_A)}{P(x_V, x_A | C)}, \quad (10)$$

where  $P(C = 1 | x_V, x_A) = \pi_C$  (the posterior mixing proportion).

Natarajan *et al.* [21] investigated these issues further. In particular, they showed that human performance on these types of experiments could be better modeled by replacing the Gaussian distributions by a more robust-alternative. It is well-known that Gaussian distributions are non-robust because the tails of their distributions fall off rapidly which gives very low probability to rare events. Hence in many real-world applications distributions with longer tails are preferred. Following this reasoning Natarajan *et al.* assumed that the observations  $x_A, x_V$  were generated by distributions with longer tails. More precisely, they assumed that the data is distributed by a mixture of a Gaussian distribution (as in the models above) and a uniform distribution which yields longer tails. More formally, they assume  $x_A \sim \pi N(x_A : s_A, \sigma_A^2) + \frac{(1-\pi)}{r_1}$  and  $x_V \sim \pi N(x_V : s_V, \sigma_V^2) + \frac{(1-\pi)}{r_1}$  where  $\pi$  is a mixing proportion and  $U(x) = 1/r_1$  is a uniform distribution defined over the range  $r_1$ .

### C. Classes of Probability Models

NEEDS REVISION TO FIT INTO THIS SECTION. MUST EXPRESS THE GENERALITY OF THE PROBABILISTIC APPROACH. THE FACT THAT NODES MOGHT REPRESENT GROUPS OF NEURONS AND CONCEPTS LIKE SURFACE.

We now briefly discuss more general classes of probability models. Some of these are illustrated in figure (6). For a general introduction to these models in cognitive science see [8]

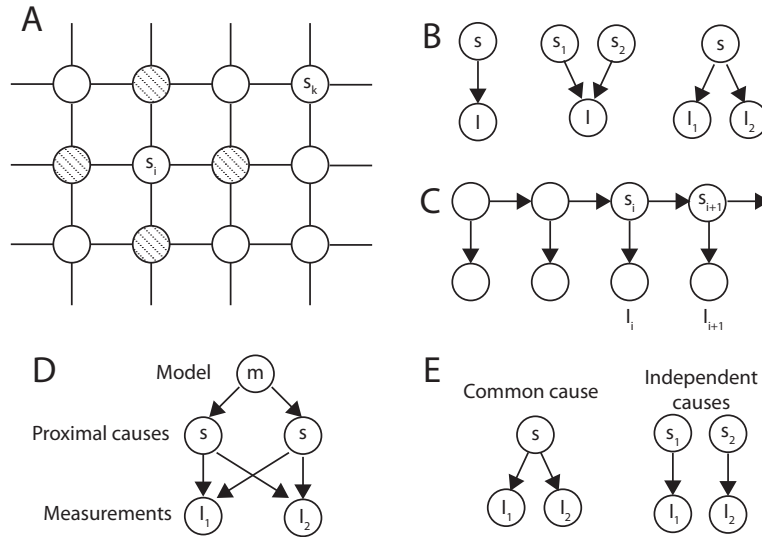


Fig. 6. Probability models defined on graphs. A specifies an undirected graphical model, which will be used for models involving spatial processing in sections (??,??). B,C,D, E show directed graphical models, which will be illustrated in later sections. Models B,D,E will be used for cue combination in section (I). Model C is used for integrating motion over time in subsection (??).

Graphical models are defined on  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ . Here  $\mathcal{V}$  denotes the set of graph nodes. State variables  $z_\mu$  are defined on the nodes  $\mu \in \mathcal{V}$ . A probability distribution  $P(\vec{z}) = P(\{z_\mu : \mu \in \mathcal{V}\})$  is defined over the graph variables.

Broadly speaking there are two types of graphical models (although hybrids exist). For undirected graphical models, see figure (6)(A), the probability distribution is constructed using potential functions



defined over maximal cliques  $c \in \mathcal{C}$ . A maximal clique is a fully connected set of nodes (i.e. there is an edge between every pair of nodes in the clique), so that we cannot extend the clique to include extra nodes without losing fully connectedness. The potential is  $\phi_c(\vec{z}_c)$ , where  $\vec{z}_c$  is the state of the nodes in the clique (e.g.,  $(z_\mu, z_\nu)$  if the clique only contains two nodes  $\mu, \nu$ ). The distribution is expressed in terms of a Gibbs distribution:  $P(\vec{z}) = \frac{1}{Z} \exp\{\sum_{c \in \mathcal{C}} \phi_c(\vec{z}_c)\}$ . This graph obeys the markov property:  $P(z_\mu | \vec{z}_\mu) = P(z_\mu | \{z_\nu : (\nu, \mu) \in \mathcal{E}\})$ . The last few subsections give examples of clique potentials, such as  $\sum_{ij} \theta_{ij} z_i, z_j$ .

In *directed graphical models*, the graph edges  $\mathcal{E}$  have direction, see figure (6)(B,C,D,E). For any node  $\mu \in \mathcal{V}$ , the set of parent nodes  $pa(\mu)$  are the set of all nodes  $\nu \in \mathcal{V}$  such that  $(\mu, \nu) \in \mathcal{E}$ , where  $(\mu, \nu)$  means that there is an edge between nodes  $\mu$  and  $\nu$  pointing to node  $\mu$ . This gives a local *markov property* – the conditional distribution  $P(z_\mu | \vec{z}_\mu) = P(z_\mu | \vec{z}_{pa(\mu)})$ , so the state of  $z_\mu$  is only directly influenced by the state of its parents (note  $\vec{z}_\mu$  denotes the states of all nodes except for node  $\mu$ ). The probabilistic models for divisive normalization are examples of directed graphical models.

We emphasize that many computational models of vision, and other aspects of intelligence, can be expressed in terms of this probabilistic formulation. This may serve as an intermediate computational level between models of neural circuits and of behavior. But how can models of these type be implemented by real neurons? This section has given examples showing that some probability models map nicely onto simple neural net models. But this mapping may be too simplistic and may also be impractical for more complicated probability models. This has motivated the study of how populations of neurons may be able to encode and process probabilities [13], [18].

## REFERENCES

- [1] A. Blake and H. B.ülthoff. Does the brain know the physics of specular reflection? *Nature*, 343(6254):165–168, 1990.
- [2] H. H. Bulthoff and H. A. Mallot. Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A*, 5(10):1749–1758, Oct. 1988.
- [3] K. Cheng, S. J. Shettleworth, J. Huttenlocher, and J. J. Rieser. Bayesian integration of spatial information. *Psychological Bulletin*, 133(4):625–637, July 2007.
- [4] J. J. Clark and A. L. Yuille. *Data Fusion for Sensory Information Processing Systems*. Springer, 1990.
- [5] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, Jan. 2002.
- [6] M. Gori, M. Del Viva, G. Sandini, and D. C. Burr. Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, 18(9):694–698, May 2008.
- [7] C. S. Green, C. Benson, D. Kersten, and P. Schrater. Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences*, 107(37):16401–16406, Sept. 2010.
- [8] T. Griffiths and A. Yuille. A primer on probabilistic inference. *The probabilistic mind: Prospects for Bayesian cognitive science*, pages 33–57, 2008.
- [9] D. Hubel. Evolution of ideas on the primary visual cortex, 1955–1978: A biased historical account. *Bioscience reports*, 2(7):435–469, 1982.
- [10] R. Jacobs. Optimal integration of texture and motion cues to depth. *Vision Research*, 39(21):3621–3629, Oct. 1999.
- [11] D. Kersten, H. H. Bulthoff, B. Schwartz, and K. Kurtz. Interaction between transparency and structure from motion. *Neural Computation*, 4(4):573–589, 1992.
- [12] D. Knill. Mixture models and the probabilistic structure of depth cues. *Vision Research*, 43(7):831–854, Mar. 2003.
- [13] D. C. Knill and A. Pouget. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719, Dec. 2004.
- [14] K. P. Körding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams. Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9):e943, Sept. 2007.
- [15] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35:389–412, 1995.
- [16] P. Lennie. Single units and visual cortical organization. *Perception*, 27:889–935, 1998.
- [17] M. Livingstone and D. Hubel. Anatomy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience*, 4(1):309–356, 1984.
- [18] W. J. Ma. Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research*, 50(22):2308–2319, Oct. 2010.
- [19] D. MacKay. *Information theory, inference, and learning algorithms*, 2003.
- [20] D. Marr. *Vision*. W.H. Freeman, 1982.
- [21] R. Natarajan, I. Murray, L. Shams, and R. Zemel. Characterizing response behavior in multisensory perception with conflicting cues. In *Advances in neural information processing systems NIPS*, 2008.
- [22] L. Shams. Probability Matching as a Computational Strategy Used in Perception. *PLoS Computational Biology*, 6(8):e1000871, Aug. 2010.
- [23] M. T. Wallace, G. E. Roberson, W. D. Hairston, B. E. Stein, J. W. Vaughan, and J. A. Schirillo. Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), Apr. 2004.



- [24] A. Yuille and H. H. Bulthoff. Bayesian decision theory and psychophysics. In D. Knill and W. Richards, editors, *Perception as Bayesian Inference*, page 123. Cambridge University Press, 1996.
- [25] A. L. Yuille and H. H. Bülthoff. Bayesian decision theory and psychophysics. In D. Knill and W. Richards., editors, *Bayesian Approaches to Perception*. Cambridge University Press, 1996.