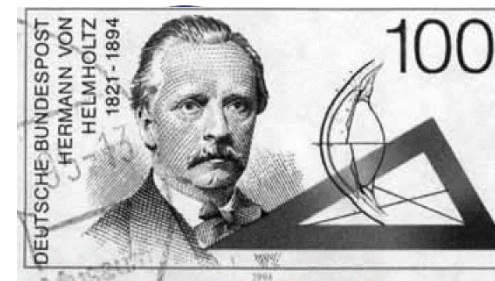


Vision as Bayesian Inference

- It is attractive to formulate Vision, and other aspect of intelligence, in terms of probabilistic inference.
- Vision is an inverse inference problem so it can be formulated as Bayesian Inference.



Vision as Inverse Inference

- Helmholtz. 1821-1894.
- More recently: inverse computer graphics.
- Vision must invert the forward process (CG) to discover the causal factors that have generated the images.
- *But only for the tasks that we care about – attention, change blindness.*

Richard Gregory

- "Perception (vision) as hypotheses".
- Perception is not just a passive acceptance of stimuli, but an active process involving memory and other internal processes.
- Humans have internal representations – we see images when we dream, we can imagine what animals and people will do, we can hallucinate.

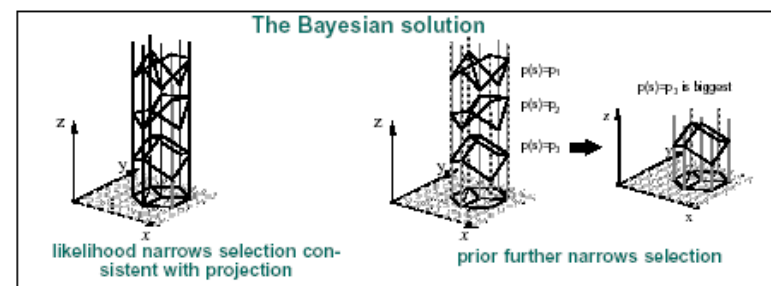
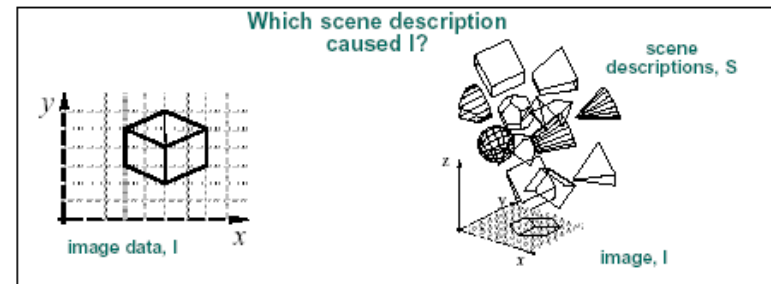
Bayesian Decision Theory

- Bayes' Theorem is a conceptual framework for inverse inference problems.
- We can infer the state S of the world from the image I using prior knowledge.
- **$P(S|I) = P(I|S)P(S)/P(I)$.**
- Rev. T. Bayes. 1702-1761
- $P(I|S)$ likelihood, $P(S)$ prior



Inverse Problems are hard

- There are an infinite number of ways that images can be formed.
- Why do we see a cube?
- The likelihood $P(I|S)$ rules out some interpretations S
- Prior $P(S)$ — cubes are more likely than other shapes consistent with the image.



Inverse inference requires priors

- Humans use prior knowledge about the world (obtained through experience). Often correct – but can fail occasionally.
- Flying carpet? Levitate?
- ***Play ball-in-box.***



Analysis by Synthesis: Mumford & Grenander

- Grenander (1960's) had proposed that vision could be formulated as pattern theory and proposed the idea of “analysis by synthesis”. This is naturally expressed in Bayesian terms.



- Mumford embraced Analysis by Synthesis and Pattern Theory.
- Analysis by Synthesis emphasizes pattern synthesis as well as pattern analysis. Bayesian inference requires you construct a prior probability model of whatever signals or situations you are modeling and you should always test your prior by sampling to see which features it models accurately and which it does not.

Mumford's Bold Hypothesis.

- Mumford (1991) boldly proposed a model for how a primate brain could perform analysis by synthesis using bottom-up and top-down processing.
- He proposed that each area of the cortex carries on its calculations with the active participation of a nucleus in the thalamus with which it is reciprocally and topographically connected. This nucleus plays the role of an 'active blackboard' on which the current best reconstruction of some aspect of the world is always displayed
- Each cortical area maintains and updates the organism's knowledge of a specific aspect of the world, ranging from low level raw data to high level abstract representations, and involving interpreting stimuli and generating actions.
- It draws on multiple sources of expertise, learned from experience, creating multiple, often conflicting, hypotheses which are integrated by the action of the thalamic neurons and then sent back to the standard input layer of the cortex.
- .

Mumford's bold hypothesis for the architecture of the neocortex.

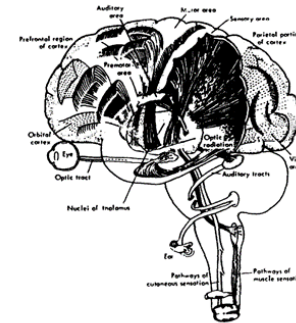


Fig. 2. The location of the thalamus within the cortex (from Luria 1969)

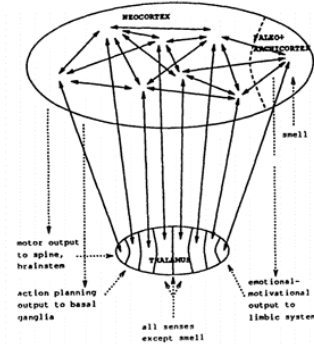


Fig. 3. Simplified schematic of cortical connections

- The higher areas of the neocortex attempts to fit its abstractions to the data it receives from lower areas by sending back to them from its deep pyramidal cells a template reconstruction best fitting the lower level view.
- The lower areas attempts to reconcile the reconstruction of its view that it receives from higher areas with what it knows, sending back from its superficial pyramidal cells the features in its data which are not predicted by the higher area.
- The whole calculation is done with all areas working simultaneously, but with order imposed by synchronous activity in the various top-down, bottom-up loops.
- *Neuroscience experiments give increasing support for top-down models and maybe for analysis by synthesis.*

Vision as Bayesian Inference: Yuille & Kersten

- A. Yuille & D. Kersten. Trends in Cognitive Science. 2006.

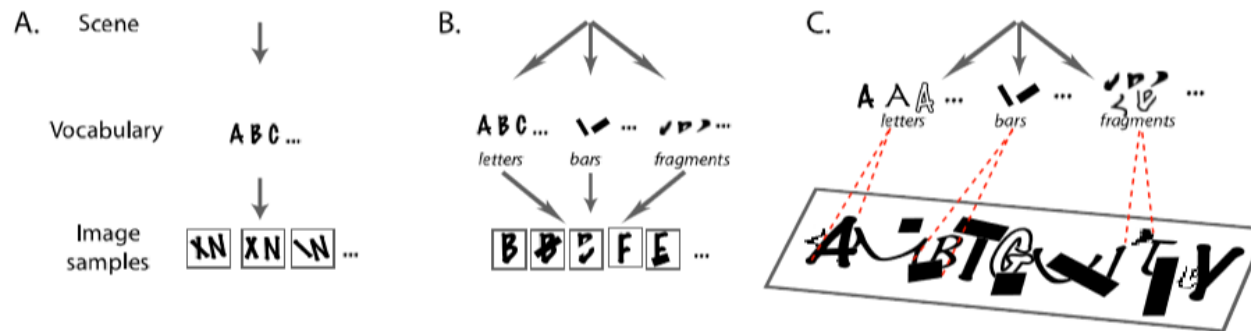


Figure 1: Left Panel (A). A simple vocabulary for generating the image. There is little, or no, ambiguity in interpreting images. At worst, the letter *X* may be confused with a slanted *I* partially occluding a vertical *I*. Centre Panel (B). A richer vocabulary. A given cause, such as a particular letter, can be manifest in many different images. But there are now multiple ways to generate identical images, see text. Right Panel (C). The richer the vocabulary, the greater the image ambiguity, and the harder it is to interpret the image. This leads to a formidable inference problem.

Vision as Bayesian Inference: Yuille & Kersten

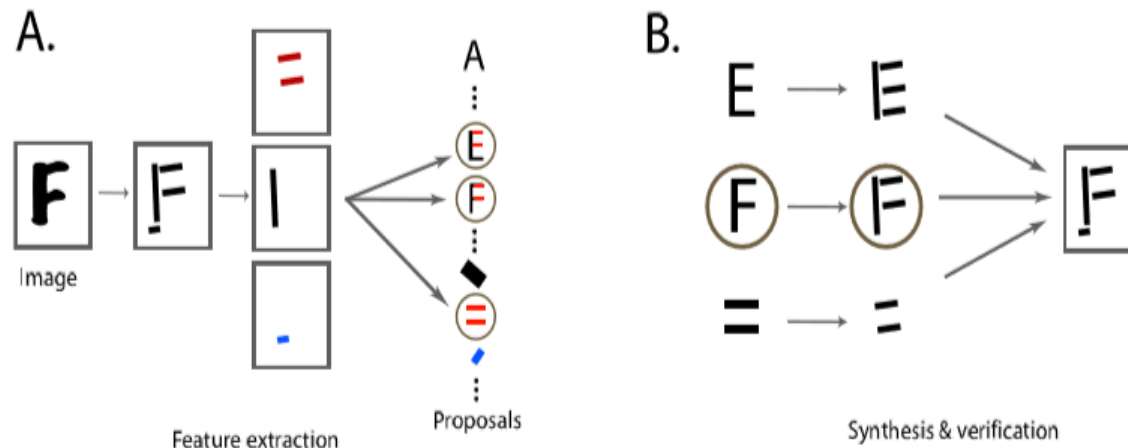


Figure 2: Analysis by synthesis. A. Low-level processing (left panel) can extract edge features, such as bars, and use conjunctions of these features to make bottom-up proposals to access the higher-level models of objects. B. The high-level objects access the image top-down to validate or reject the bottom-up proposals (right panel). In this example, the low-level cues propose that the image can be interpreted as an *E*, an *F*, or a set of parallel bars. But interpreting it as an *F* explains almost all the features in the image and is preferred.

Vision as Bayesian Inference: Yuille & Kersten

- Adding more realism building on work by Z. Tu & S.C. Zhu 2002, Z. Tu et al. 2006.

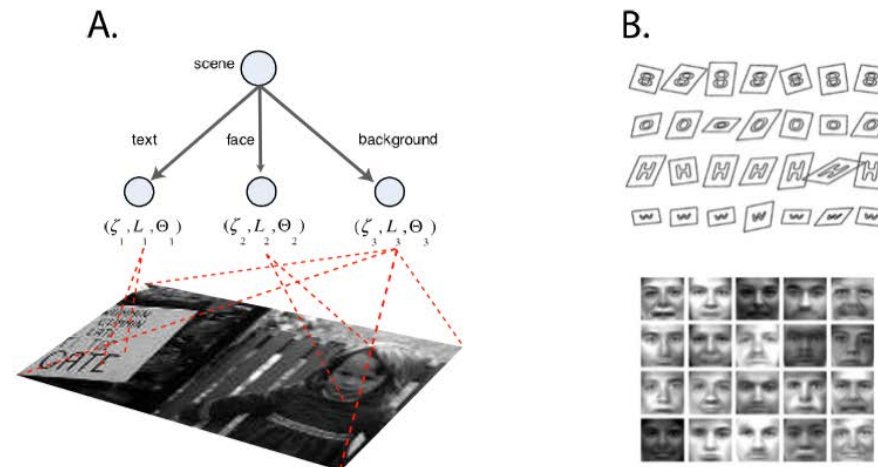


Figure 3: A. The image is generated (left panel) by a probabilistic context free grammar shown by a two layer graph with nodes with properties (ζ, l, θ) corresponding to regions L_i in the image. B. The right panel shows samples from the face model and the letter model – i.e. from $p(I_{R(L)}|\zeta, L, \Theta)$.

Vision as Bayesian Inference

- This is ***Bayes-in-the-Big***. It is a highly ambitious research program which is conceptually very attractive.
- Challenges include learning probability models to generate images of all the classes of objects, and background “stuff”, that occur in real images.
- Efficient inference algorithms that can take an image as input and rapidly search over all the possible world configurations (of objects and background) than can have caused it.