# Applications of Bayesian Networks

- modelling human multimodal perception
  - human sensor data fusion
  - top down influences in human perception
- multimodal human-computer interaction

# Human Sensor Data Fusion

- two general strategies (Ernst and Bülthoff, 2004)
  - sensory combination: maximize information delivered from the different sensory modalities
  - sensory integration: reduce the variance in the sensory estimate to increase its reliability

# Sensor Data Fusion

- sensory integration has to produce a coherent percept
- Which modality is the dominating one?
  - visual capture: e.g. vision dominates haptic perception
  - auditory capture: e.g. number of auditory beeps vs. number of visual flashes
- modality precision, modality appropriateness, estimate precision: the most precise modality wins

# Sensor Data Fusion

- two possible explanations:
  - maximum likelihood estimation: weighted sum of the individual estimates
    - all cues contribute to the percept
  - cue switching:
    - the most precise cue takes over
    - the less precise cues have no influence

# Sensor Data Fusion

- maximum likelihood estimate:
  - weighted sum of the individual estimates
  - weights are proportional to their inverse variance

$$\hat{s} = \sum_i w_i \; \hat{s}_i \text{ with } \sum_i w_i = 1$$

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$$

  - most reliable unbiased estimate possible (estimate with minimal variance)
  - optimality not really required; good approximation might be good enough

# Sensor Data Fusion

- overwhelming evidence for the role of estimate precision
- weighting within modalities
  - visual depth perception: motion + disparity, texture + disparity
  - visual perception of slant
  - visual perception of distance
  - haptic shape perception: force + position
- cross modal weighting:
  - vision + audition
  - vision + haptic
  - vision + proprioception

# Sensor Data Fusion

- no conclusive evidence for the reliability hypothesis so far
- How to estimate the variance of a stimulus?
  - requires an independence assumption
  - difficult to achieve in a unimodal task
  - cues within one modality are correlated
  - $\rightarrow$ multi-modal experiments

# Sensor Data Fusion

- Ernst and Banks (2002): vision-haptic integration
  - modifying the visual reliability by adding noise to the visual channel
  - two extreme cases:
    - vision dominates (little noise)
    - haptics dominate (high noise)

$\rightarrow$ perception requires dynamic adjustment of weights

$\rightarrow$ nervous system has online access to sensory reliabilities

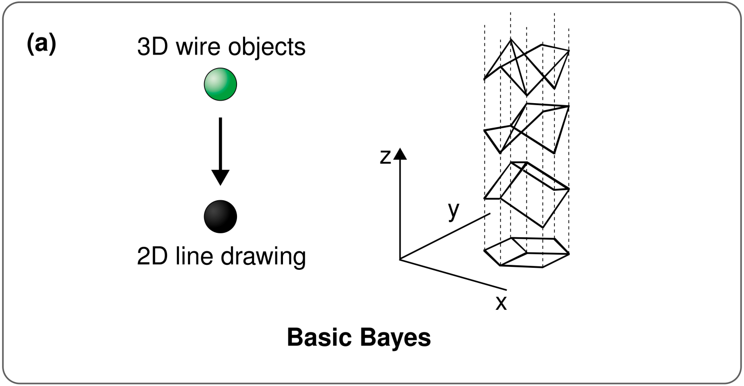
# Sensor Data Fusion

- But where do the estimates come from?
- prior experience vs. on-line estimation during perception
- on-line is more likely: observing the fluctuations of responses to a signal
    - over some period of time
    - across a population of independent neurons (population codes)

# Top Down Influence

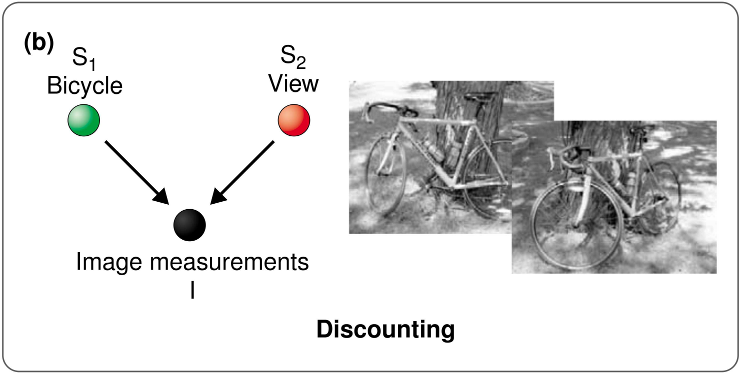
- perception is modulated by contextual factors, e.g scene or object properties
- How to model top-down influences?
  - can be captured by prior probabilities
  - prior probabilities can be integrated by means of Bayes rule
    $\rightarrow$ Bayesian reasoning

# Top Down Influence



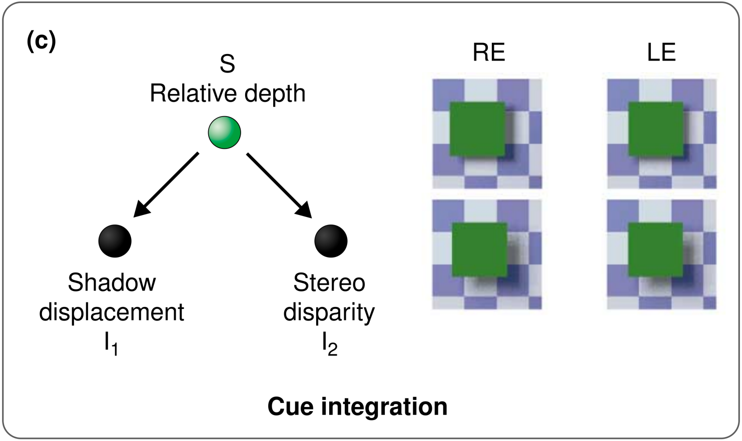

Kersten and Yuilley (2003)

# Top Down Influence



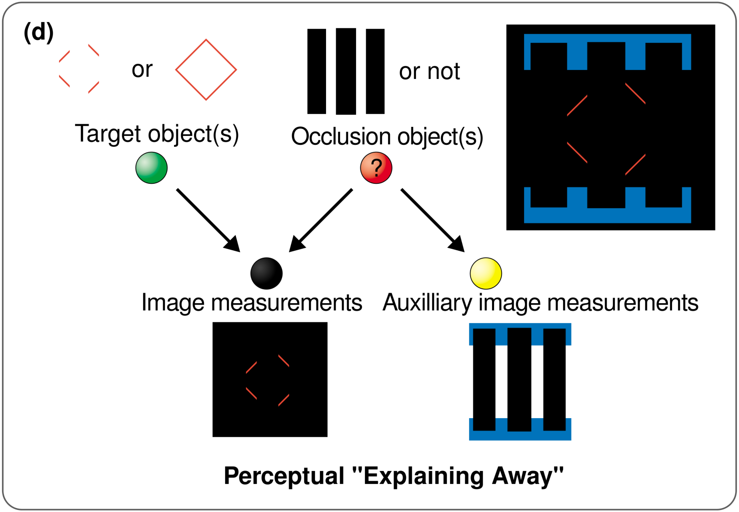

Kersten and Yuilley (2003)

# Top Down Influence




Kersten and Yuilley (2003)

# Top Down Influence




Kersten and Yuilley (2003)

# Multimodal Human-Computer Interaction

- Socher, Sagerer, Perona (2000), Wachsmuth, Sagerer (2002)
  - multi-modal human machine interaction using
    - speech
    - vision
    - (pointing gestures)

- data fusion from different reference systems
  - spatial (vision) vs. temporal (speech)
  - language based instruction: fusion on the level of concepts

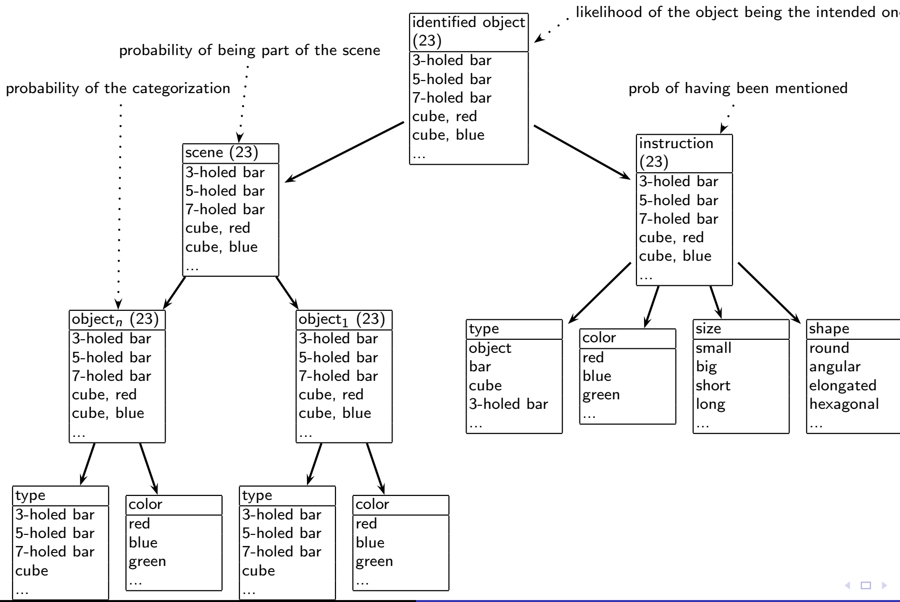
# Multimodal Human-Computer Interaction

- noisy and partial interpretation of the sensory signals
- dealing with referential uncertainty
- goal: cross modal synergy

- sensory data: properties (color) and (spatial) relationships: degree-of-membership representation (fuzzyness)

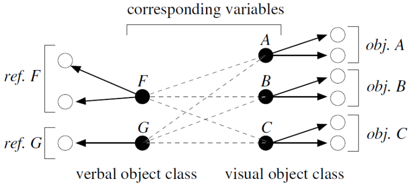
- combination using Bayesian Networks
- estimating the probabilities by means of psycholinguistic experiments
  - how do humans categorize objects and verbalize object descriptions

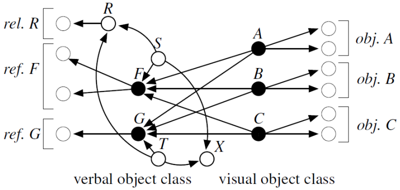
# Multimodal Human-Computer Interaction



# Multimodal Human-Computer Interaction

- more sophisticated fusion model (Wachsmuth, Sagerer 2002)
  - solution to the correspondence problem using selection variables



# Multimodal Human-Computer Interaction

- results for object identification

| | correct input | noisy speech | noisy vision | noisy input |
|---|---|---|---|---|
| recognition error rates | – | 15% | 20% | 15%+20% |
| identification rates | 0.85 | 0.81 | 0.79 | 0.76 |
| decrease of identification rates | – | 5% | 7% | 11% |

- synergy between vision and speech
- higher robustness due to redundancy between modalities