Bayesian inference and cognition

Yaara Erez & Helen Blank

Bayesian theory IG
18 February 2014

Questions

- What is the behavioural evidence for Bayesian inference as a model for **perception**?

- …for **vision**, in particular?

- …for **decision making** and **cognition**?

- …for **action** and **sensorimotor control**?

Layout

- Part 1: Bayesian inference and cognition
  - Learning and reasoning
- Part 2: Visual perception and decision making
- Discussion: so do we actually use Bayesian inference in cognition/perception/decision making?

part 1: Bayesian inference as a model for learning and reasoning

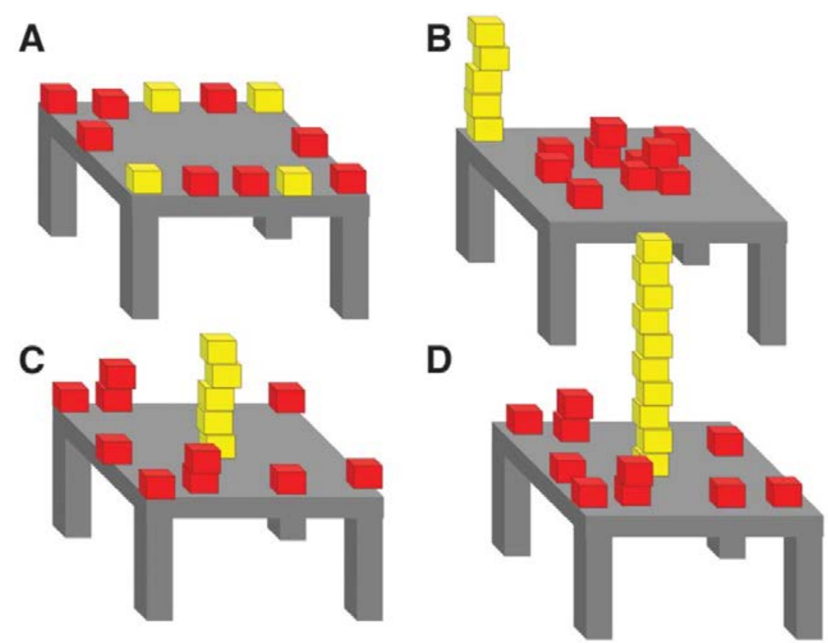
Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference

Ernő Téglás,[1,2]* Edward Vul,[3]* Vittorio Girotto,[4,5] Michel Gonzalez,[5]
Joshua B. Tenenbaum,[6]† Luca L. Bonatti[7]†

Science, 2011

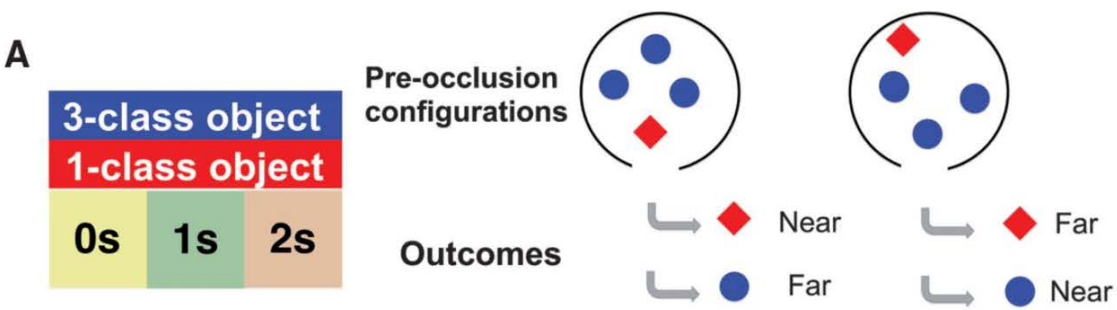
‘Pure reasoning’

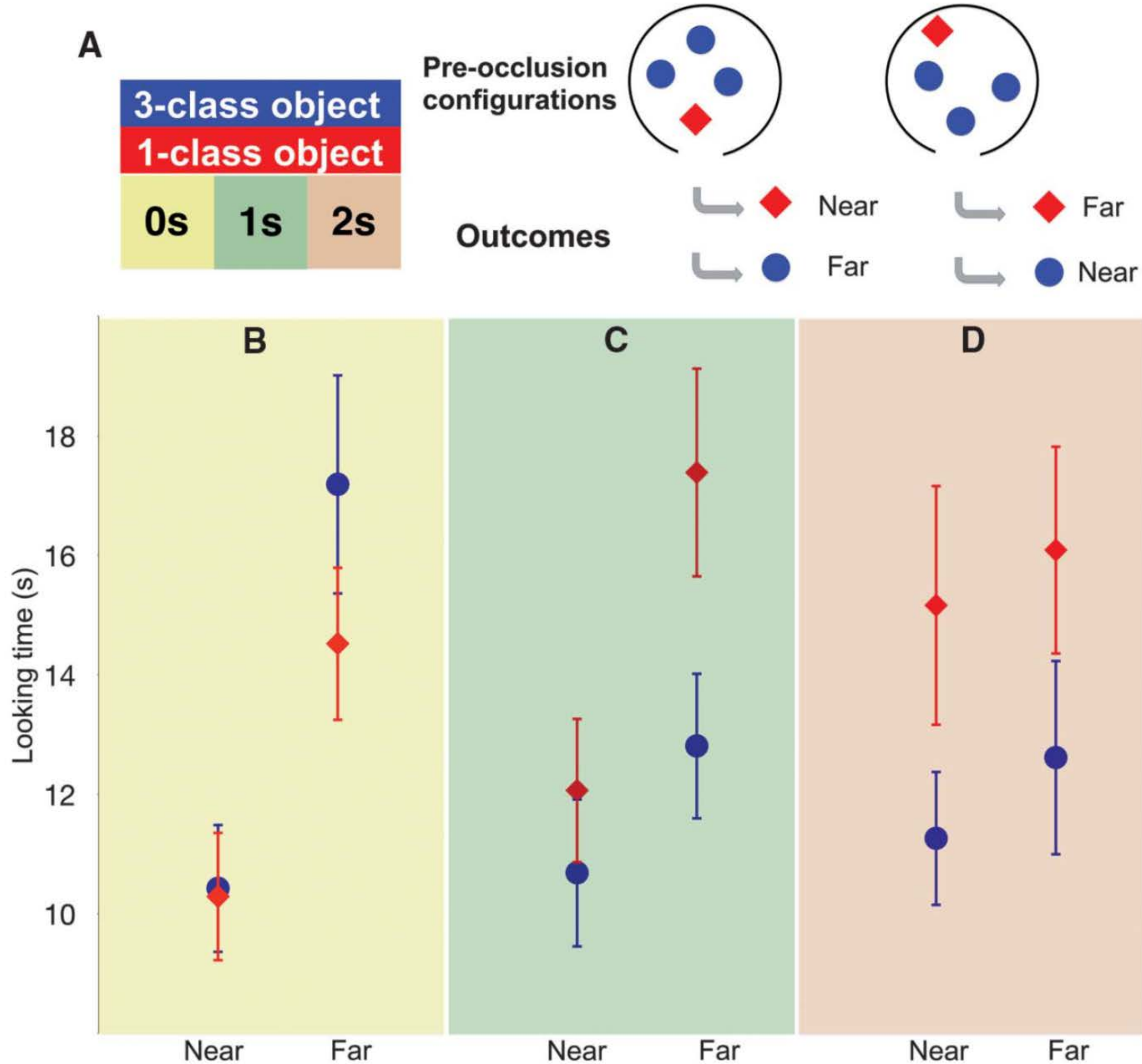
- ‘**Pure reasoning**’ – reasoning about novel situations, flexibly combining abstract knowledge and perceptual information in ‘one-shot’ intuitions to predict outcomes of events that have never been directly experienced before. **Common-sense**.
  - To distinguish from more data-driven means of forming expectations on the basis of statistical learning or finding patterns from repeated exposures.



Study objectives and experiment

- Goal: to probe the roots of 'pure reasoning' in human infants.
- Measuring **looking times** as an index of surprise.
- 12 movies, 3 factors relevant to predicting the outcome:
  - Number of objects of each type
  - Physical arrangement of objects (near/far from exit)
  - Duration of occlusion (0,1,2 s)




A
3-class object
1-class object
0s
1s
2s
Pre-occlusion configurations
Outcomes
Near
Far
Far
Near
B
C
D
Looking time (s)
18
16
14
12
10
Near
Far
Near
Far
Near
Far

A Bayesian model for infants' pure reasoning

$P(S_t|S_{t-1})$ **Prior of object dynamics** – how the state *S* of the world at time *t* depends on the state at time *t-1*

K hypothetical trajectories (sequences of states $S_{0,\ldots,F}$)

A Bayesian model for infants' pure reasoning

Probability of **final outcome**, given the observed data $D_{0,...,F-1}$

K hypothetical trajectories (sequences of states $S_{0,...,F}$)

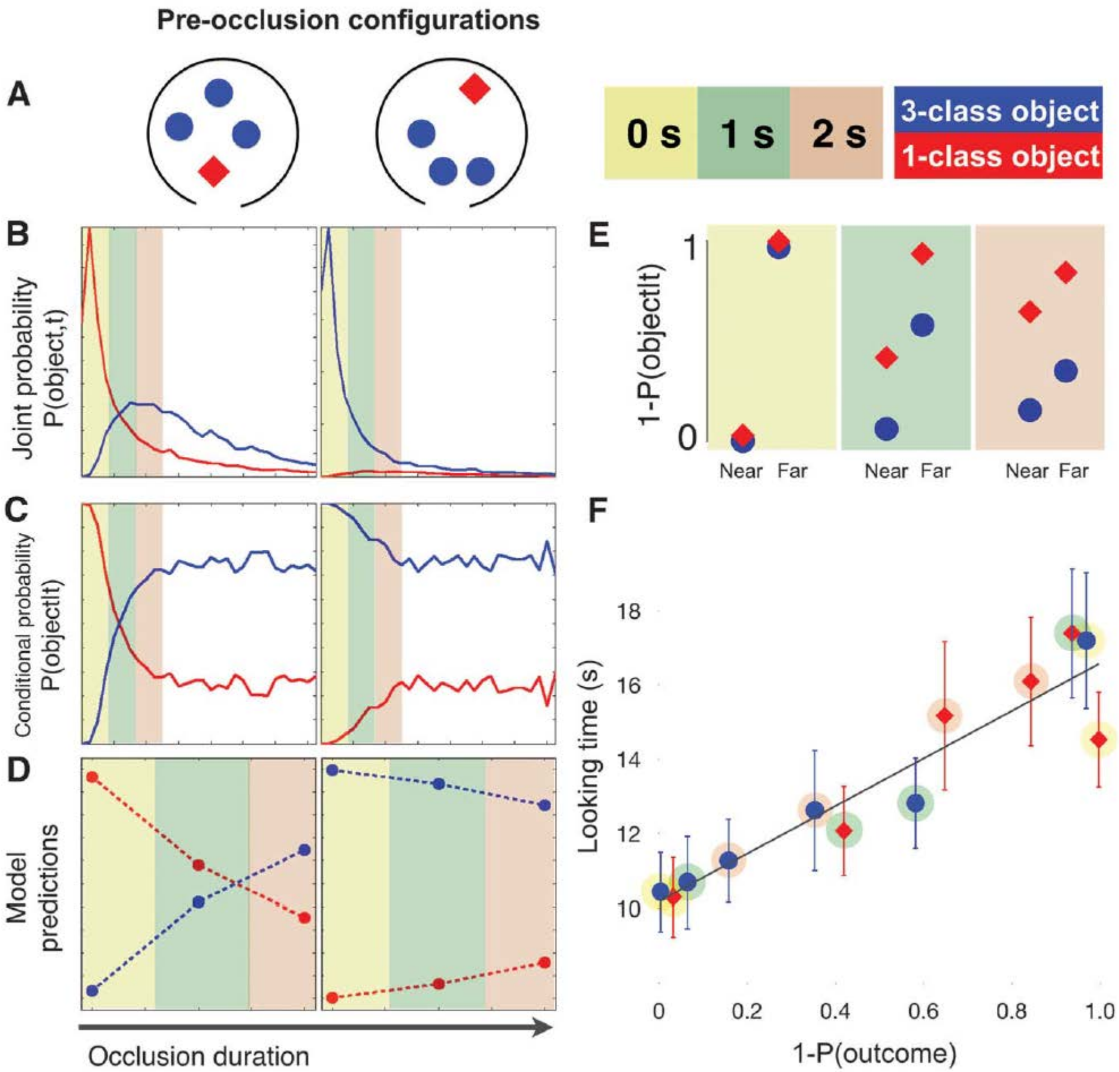
The probability of a final outcome given the state under the *k*-th trajectory

$$P(D_F | D_{0,\ldots,F-1}) \propto \sum_{k=1}^{K} P(D_F | S_F^k) \times \prod_{t=1}^{F} P(D_{t-1} | S_{t-1}^k) P(S_t^k | S_{t-1}^k)$$

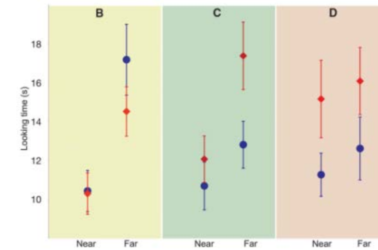
An observed outcome is **expected** insofar as many predicted future **trajectories** are **consistent** with it or unexpected if it is consistent with few predicted trajectories.

Likelihood - how well the *k*-th hypothesis fits the observed data at time point *t*

How probable is the state at time point *t* under the *k*-th hypothesis, following the **prior** of object dynamics (i.e. previous state)



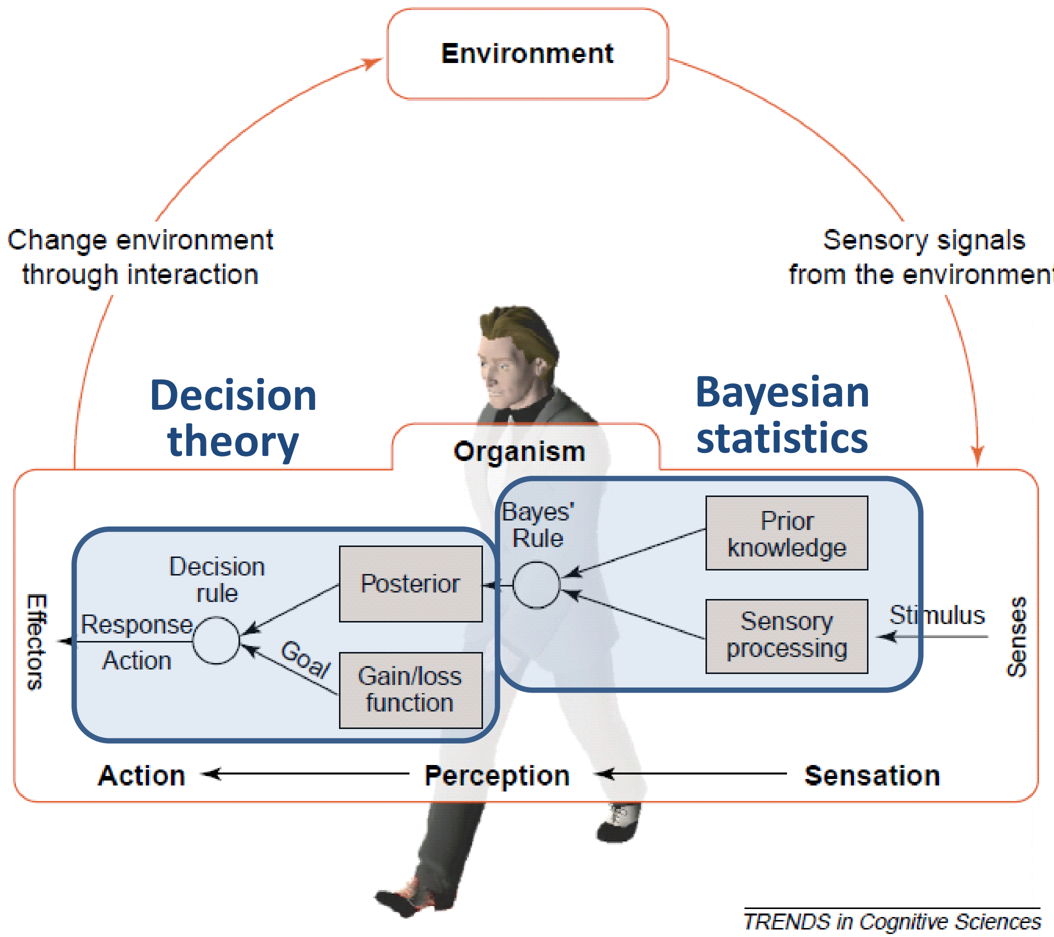
Experimental results



88% explained variance ($r=0.94$, $p<0.0001$)

- The best linear combination of the 3 factors explains 61%, and each of the factors explains significantly less than the Bayesian model.
- Resources limit / processing capacity: it is unlikely that all K trajectories are considered by humans. Model performance was similar when only 1 or 2 trajectories were used.

part 2: What is the behavioural evidence for Bayesian inference as a model for **decision making**?



Ernst & Bülthoff, 2004

Bayesian statistics

$$\overbrace{\mathrm{P(state|sensory\ input)}}^{\text{Posterior}} = \frac{\overbrace{P\,(\text{sensory input}|\text{state})}^{\text{Likelihood}}\ \overbrace{P(\text{state})}^{\text{Prior}}}{P(\text{sensory input})}$$

Decision theory

$$\sum_{\text{states}} \mathrm{L(action,state)P(state|sensory\ input)}$$

Example for loss function

Example: ***Should you eat the Fugu?***

- **probability:** 1 person in 10,000 becomes ill from the dish
  - probability of illness if you eat the Fugu of 0.0001
- **loss function:**
  - suppose you regard the loss of becoming ill from Fugu as 5,000
  - the loss of eating good Fugu as -1 (negative loss = pleasure)

- L(eat, bad Fugu) P(bad Fugu) + L(eat, good Fugu) P(good Fugu)
- which is 5,000 x 0.0001 - 1 x (1-0.0001) = 0.5-0.9999 = -0.4999

  → ***Eat the Fugu!***

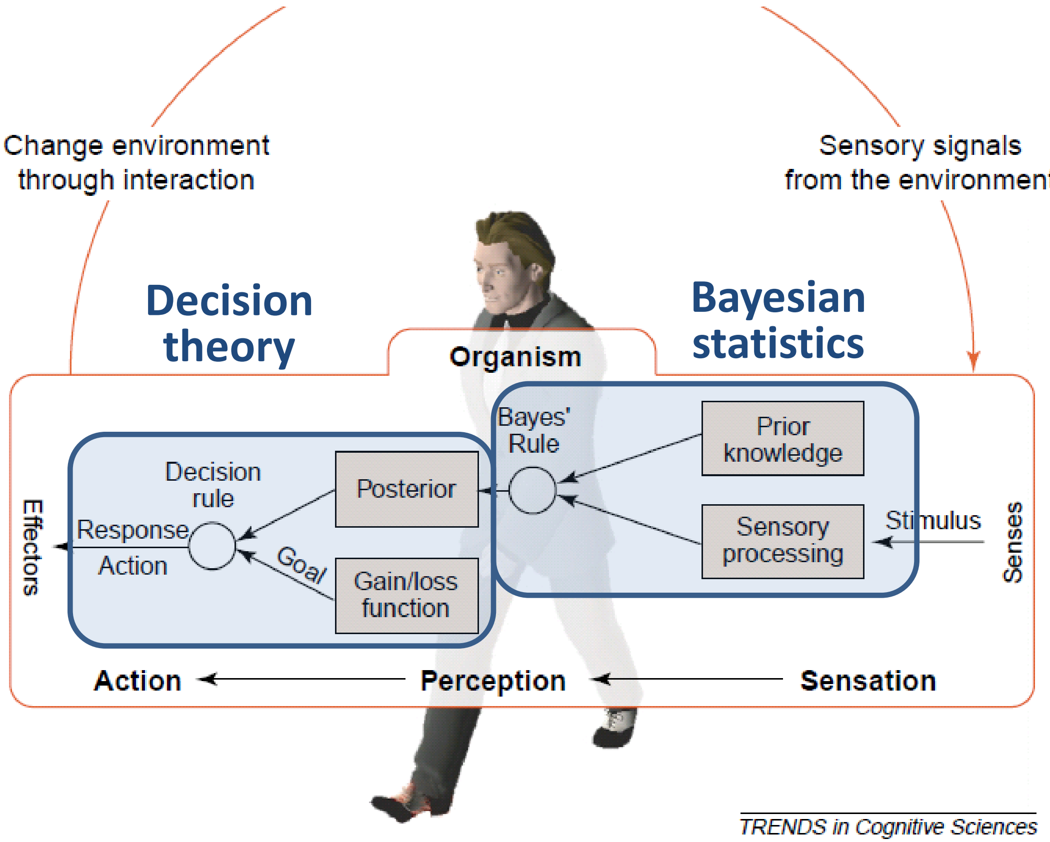
Optimal reward harvesting in complex perceptual environments

Vidhya Navalpakkam[a,1], Christof Koch[a,b], Antonio Rangel[c], and Pietro Perona[b]

Divisions of [a]Biology, [b]Engineering and Applied Science, and [c]Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125

Edited by Anne Treisman, Princeton University, Princeton, NJ, and approved January 22, 2010 (received for review October 16, 2009)

-> How does the brain combine sensory evidence and value?

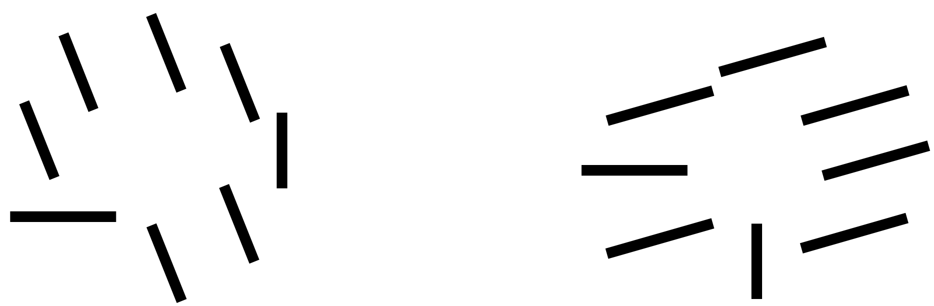


Ernst & Bülthoff, 2004

Combination of sensory evidence and value during decision making

2 targets: 

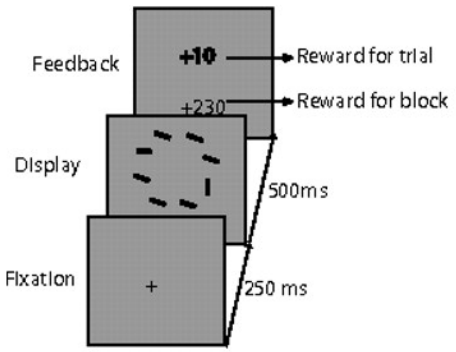
Manipulations

1. saliency: 

2. value: e.g. 20 points 10 points

Combination of sensory evidence and value during decision making

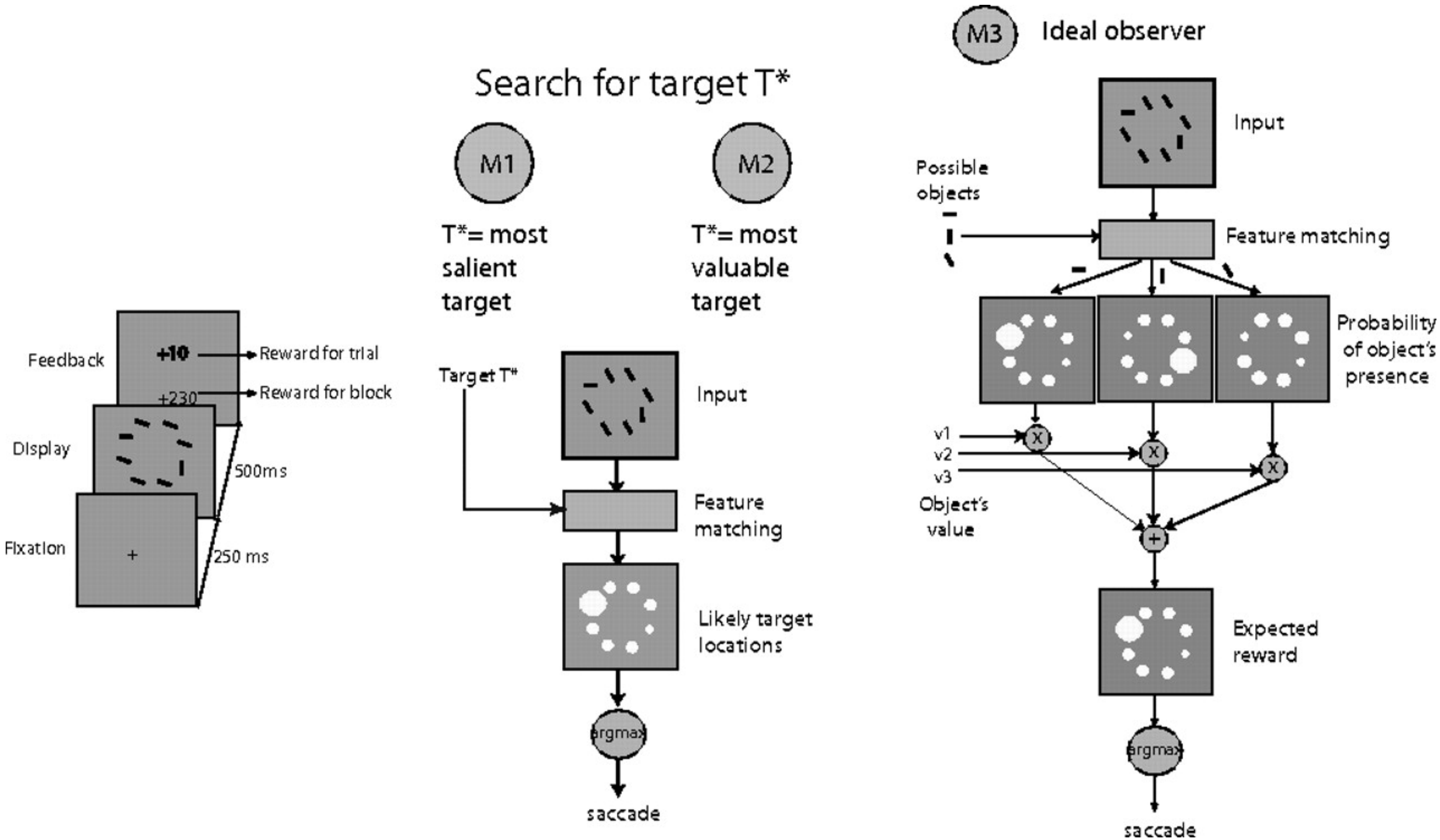
Subjects earned a reward for fixating a target for at least 100 ms.



At the beginning of each block: information about value of targets (e.g., H = 20 points, V = 10 points) + training.

Navalpakkam, et al. PNAS 2010

Combination of sensory evidence and value during decision making



most salient or most valuable target? **or ideal combination?**

Navalpakkam, et al. PNAS 2010

Combination of sensory evidence and value during decision making

1. Display consists of *n* stimuli (2 targets, H and V, as well as *n*-2 distractors, D).
2. Probability that any stimulus occupies any position is the same.
3. T_x = the stimulus feature
4. a_x = the estimate of the stimulus feature at location *x*
5. $\overrightarrow{a}$ = the resulting vector of estimates at eight locations in the display

posterior probability

$$P(T_x|\overrightarrow{a}) = \frac{P(\overrightarrow{a}|T_x)P(T_x)}{P(\overrightarrow{a})} \quad \textbf{[1]}$$

prior probability

$$P(T_x) = \frac{1}{n}, \text{ if } T_x \in \{\mathrm{H}, \mathrm{V}\}; \frac{n-2}{n}, \text{otherwise} \quad \textbf{[2]}$$

Combination of sensory evidence and value during decision making

likelihood term

$$P(\overrightarrow{a}|T_x = H) = P(a_x|T_x = H) \sum_{y \neq x} \left(P(T_y = V) P(a_y|T_y = V) \prod_{z \neq x,y} P(a_z|T_z = D) \right) \quad [3]$$

$$P(\overrightarrow{a}|T_x = V) = P(a_x|T_x = V) \sum_{y \neq x} \left(P(T_y = H) P(a_y|T_y = H) \prod_{z \neq x,y} P(a_z|T_z = D) \right) \quad [4]$$

$$P(\overrightarrow{a}|T_x = D) = P(a_x|T_x = D) \sum_{y \neq x} \left(P(T_y = H) P(a_y|T_y = H) \sum_{z \neq x,y} \left(P(T_z = V) P(a_z|T_z = V) \prod_{w \neq x,y,z} P(a_w|T_w = D) \right) \right) \quad [5]$$

Combination of sensory evidence and value during decision making

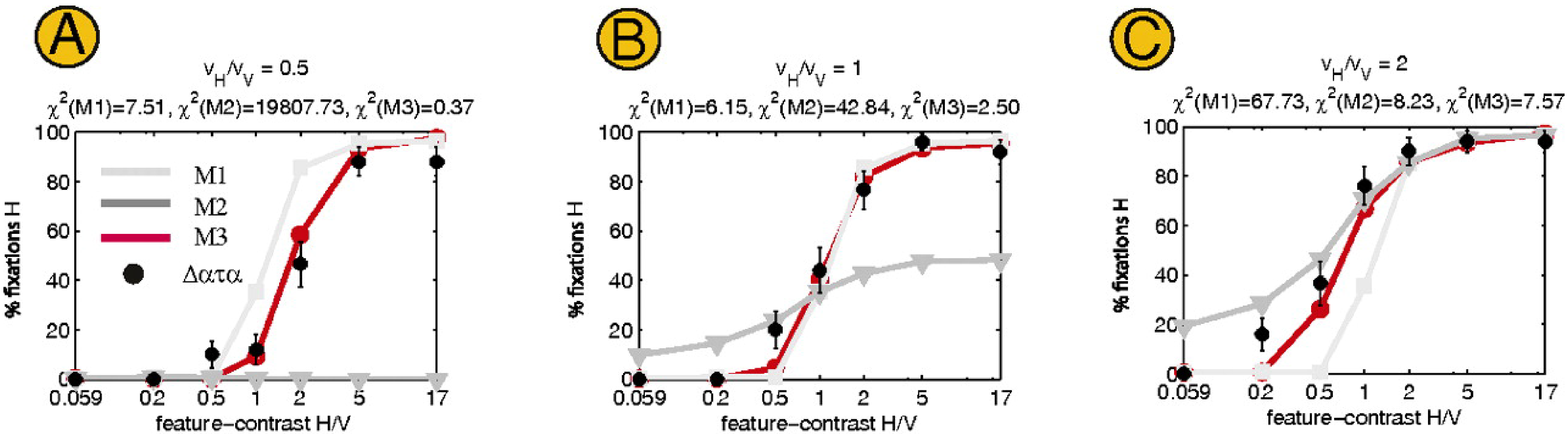
- **M1**: decision is dominated by visual properties:
  e.g., V = the more salient target; according to M1, subjects will choose the location *x*, where $P(T_x = V|\overrightarrow{a})$ is maximal

- **M2**: decision is dominated by economic properties of targets:
  e.g., if the most valuable target is H, then subjects, according to M2, will choose the location *x* where $P(T_x = H|\overrightarrow{a})$ is maximal

- **M3**: subjects compute the expected reward at every location *x* and then choose the location associated with the highest expected reward

$$E[R_x] = \sum_{i=\{D,H,V\}} v_i P(T_x = i|\overrightarrow{a}) \qquad \textbf{[6]}$$

Combination of sensory evidence and value during decision making



M3 explains data best

Navalpakkam, et al. PNAS 2010

Combination of sensory evidence and value during decision making

- Why are model parameters fitted on data of panel C?

- What about alternative models of combining seonsory evidence and value?

- Is this *optimal Bayesian* combination of evidence and value only true for *learned probabilites*?

Do we use Bayesian statistics to make decisions in general?

Say a doctor performs a test that is 99% accurate, and your test is positive for the disease.

However, the incidence of the disease is 1/10,000.

-> Your actual chance of having the disease is 1%, because the population of healthy people is so much larger than the disease.

Do we use Bayesian statistics to make decisions in general?

P(disease |pos Test) =

P(disease)P(pos Test| disease) / P(pos Test)

P(pos Test| disease) = 0.99, P(disease) = 0.0001,

P(pos Test) =

P(pos Test| disease)P(disease) +

P(pos Test| not disease)P(no disease)

= 0.99*0.0001+0.01*0.9999

Thank you!

Do we use Bayesian statistics to make decisions in general?

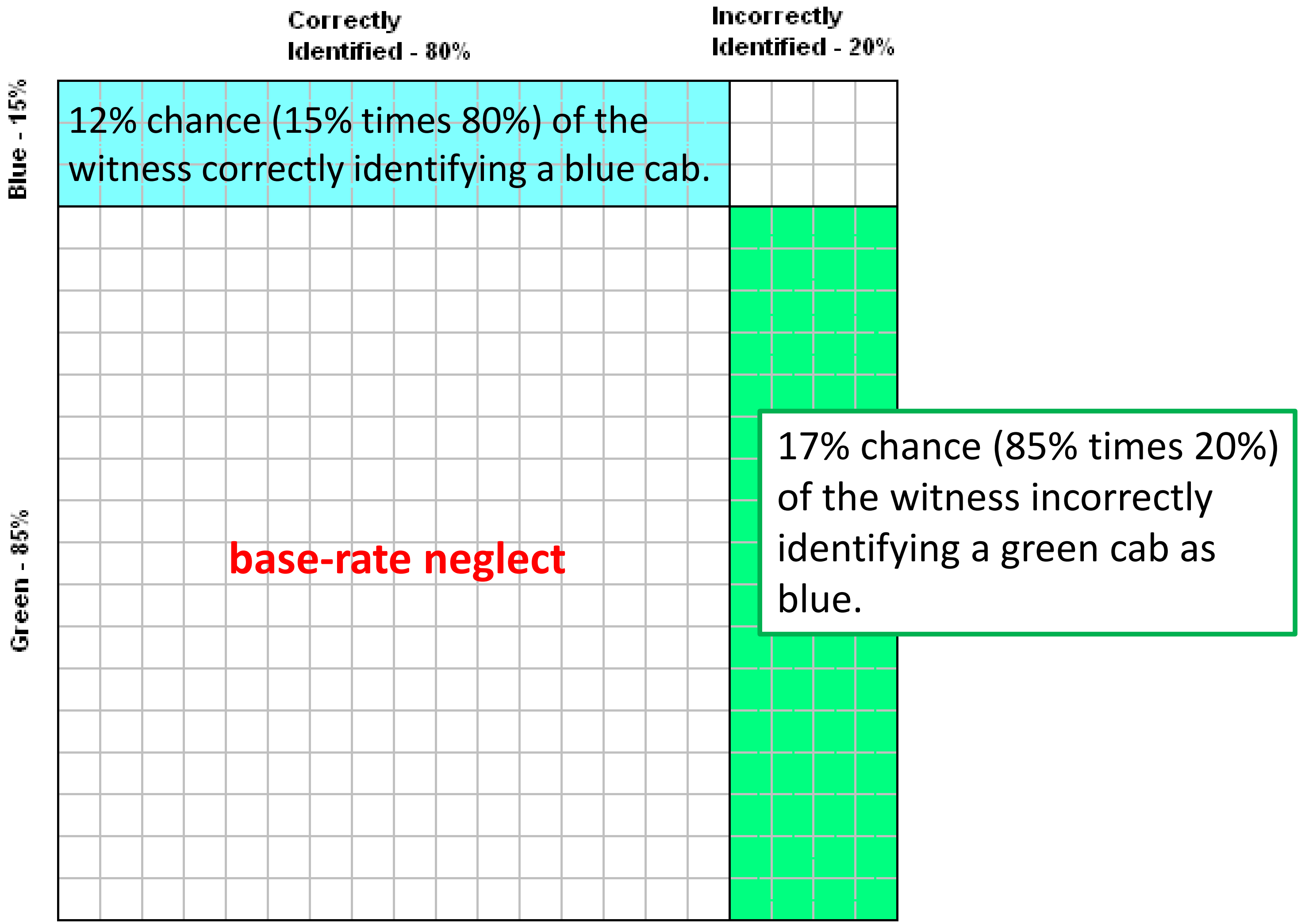
Tversky and Kahneman, subjects were given the following problem:

“A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. 85% of the cabs in the city are Green and 15% are Blue. A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time and failed 20% of the time.

- What is the probability that the cab involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?"

- Most subjects gave probabilities over 50%, some over 80%.
- The correct answer (based on Bayes) is lower than these estimates!

Do we use Bayesian statistics to make decisions in general?



= 29% chance (12% plus 17%) the witness will identify the cab as blue

= 41% chance (12% divided by 29%) that the cab identified as blue is actually blue.