Fast and frugal trees for Decision Making

Laura Martignon, Tim Erickson,
Joachim Engel, Ulrich Hoffrage, Jan Woike
Ludwigsburg University of Education
University of Lausanne
MPI for Human Development

Decisions are often made as consequences of Classifications

Example: The physician classifies a patient as being at high risk of having an infarction → she sends the patient to the coronary care unit

Thus „Decision Making made based on data“ is often the consequence of „Classification based on data“

Some facts:

- *We classify all the time*
- *The brain is a classification machine*

  It classifies based on „cues“ or „features“ that it „puts together“
- *We are good at classifying based on „predictive“ cues:*

Example: If it has thorns then it is a rose

- *Inferences like this one are often „under uncertainty and the brain often uses „frequency estimates“ in other words „statistics“.*
- *Today we need statistical thinking for inferences, classifications and decision making.*

More facts:

A statistically literate citizenry is essential to a healthy society

Emphasis on education for risk literacy is growing

Effective pedagogical approaches draw upon:

Normative theory of inference and decision-making under uncertainty is well established

Behavioral research on perception and response to risk and uncertainty is also well established

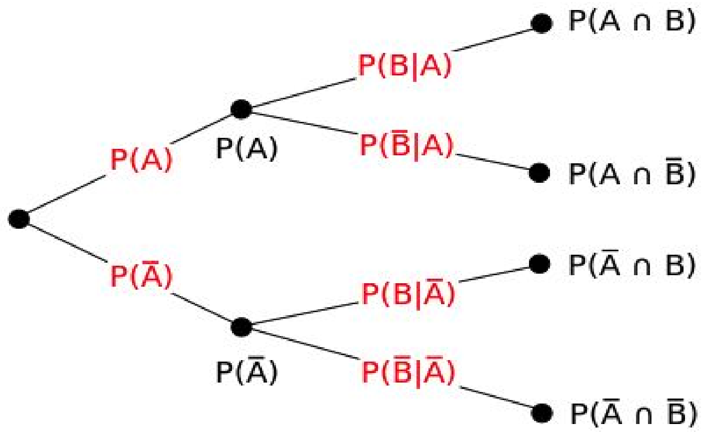
Education research on how people learn about risk and make decisions is in progress

Probability and statistics are society's primary tools for modeling and analyzing classification and decision making

… but human judgment differs systematically from probability calculations in ***some*** contexts

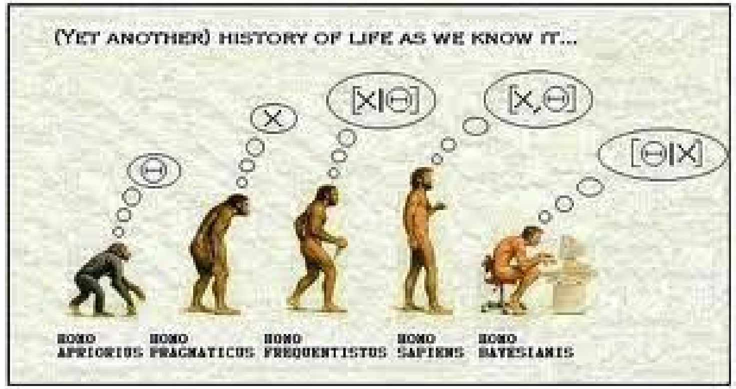
… and probability theory is often difficult for students to grasp

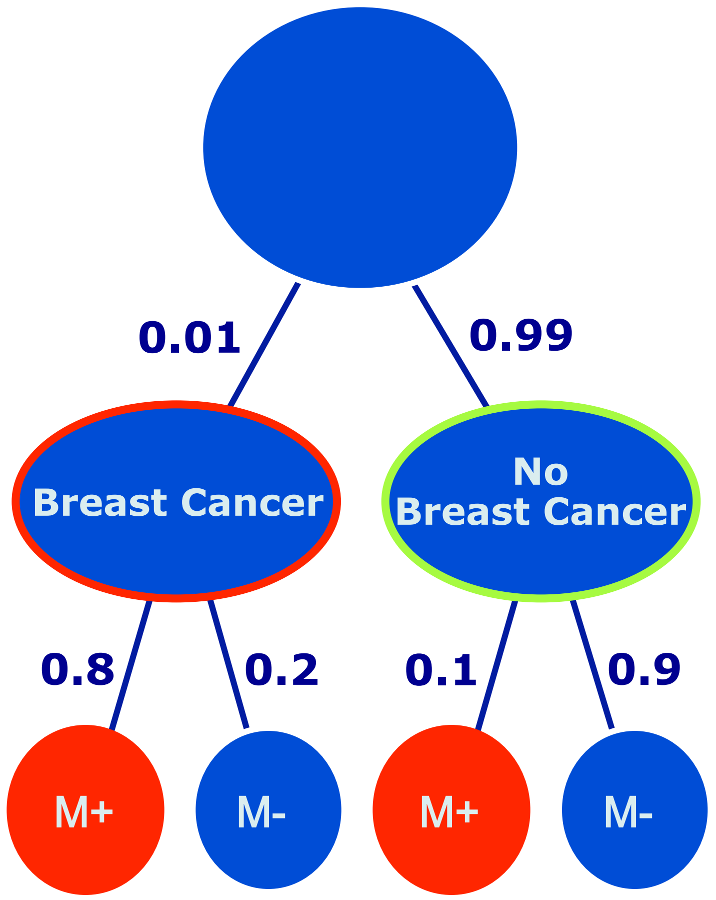
… so when and how should students encounter probability?



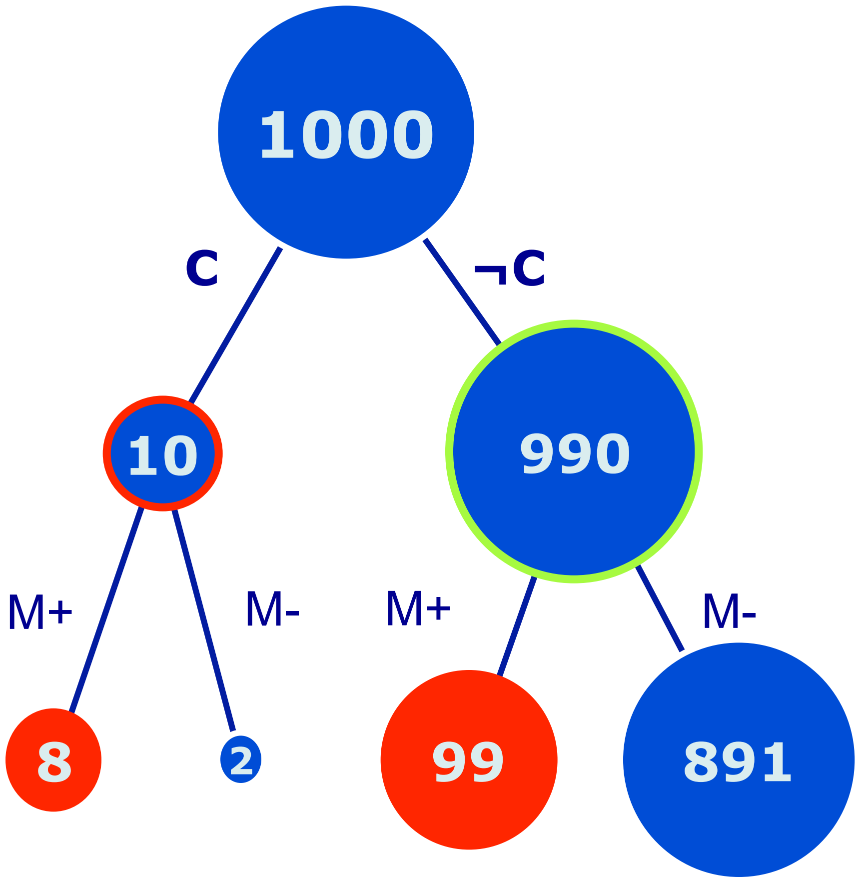
Ecological Rationality Hypothesis

- Humans have limited cognitive capacity
- Natural selection has favored approaches that provided survival advantages in environments faced by our ancestors
- In such environments, human reasoning can be as effective as normative methods
- Instructional approaches will be most effective if they build on our natural ecologically rational strategies
- Probabilistic concepts (even Bayesian inference) can be "built" or scaffolded by means of ecologically rational steps



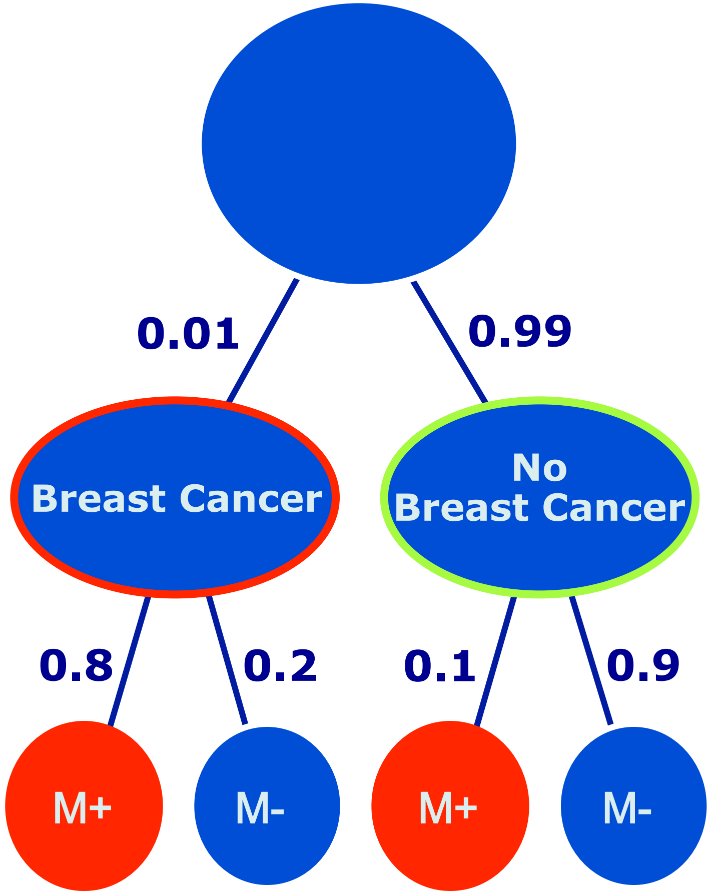


P(C|M+) = 7%

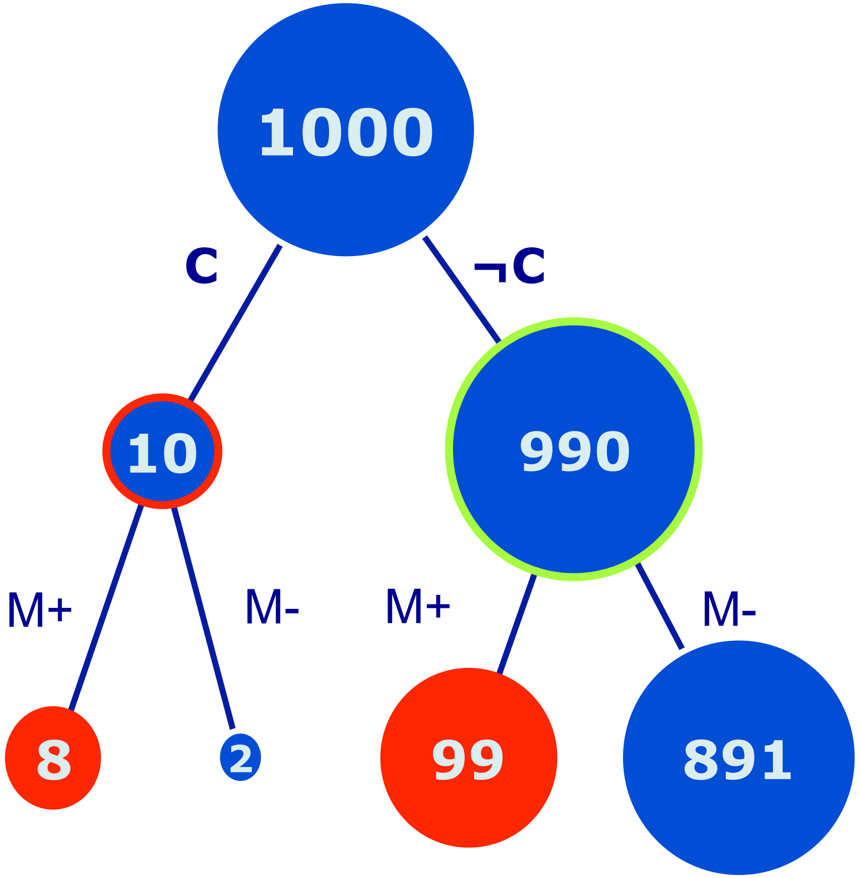


P(C|M+) = 8 of 107

Even children are good



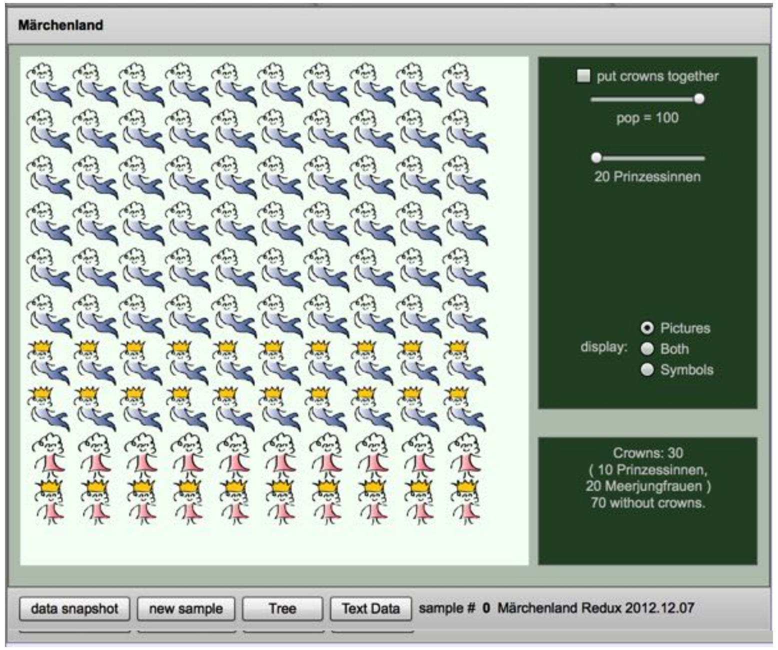
P(C|M+) = 7%

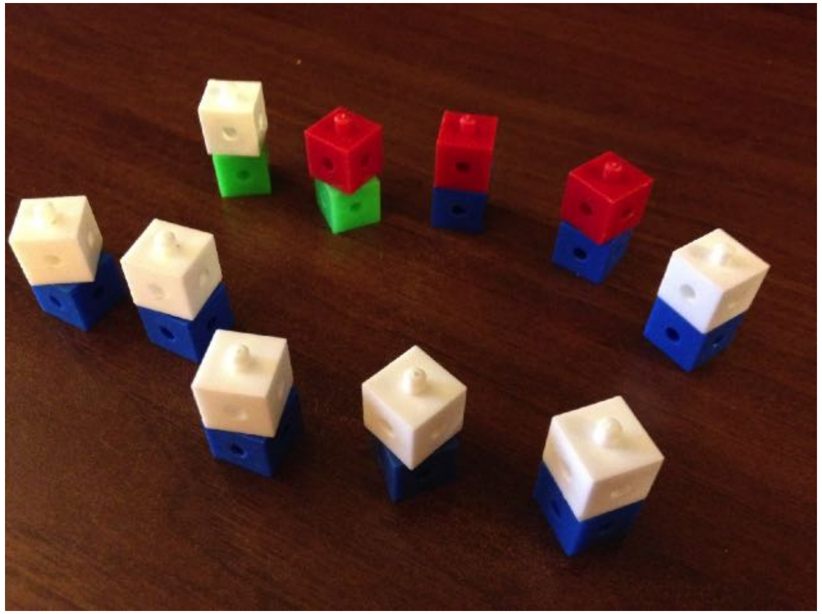


P(C|M+) = 8 of 107

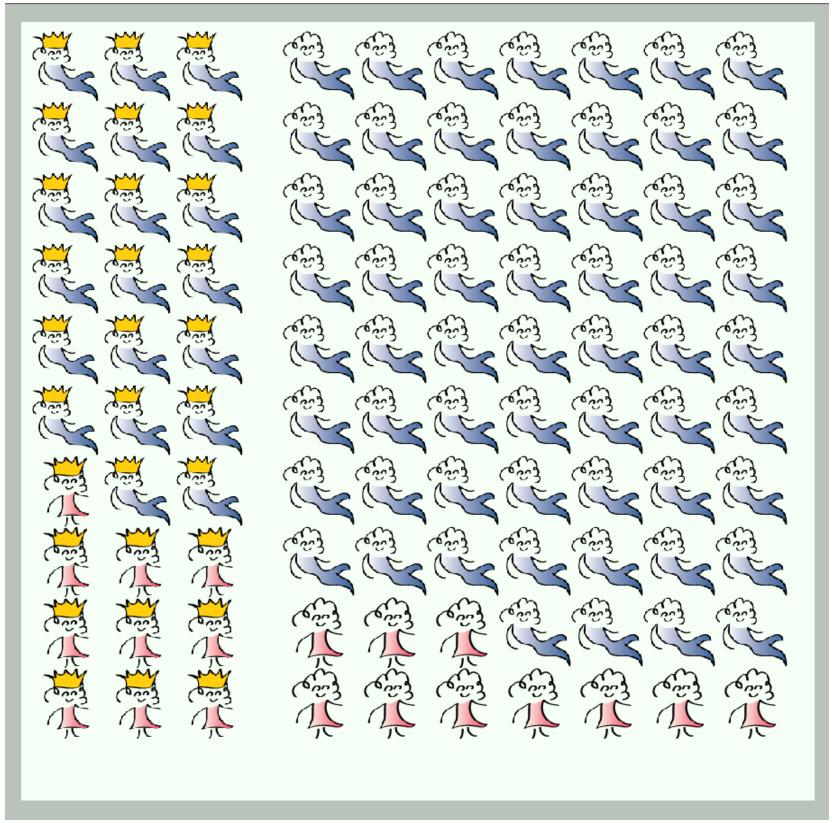
Iconic and Enactive Representations

- Tell stories
- Visualize
- Count
- Compare
- Sort
- Manipulate





Bayesian Reasoning



- Prior proportion:
  - 20 princess out of 100 Fairy Folk
- Proportions with crowns:
  - 10 out of 20 princesses have crowns
  - 20 out of 80 mermaids have crowns
- Posterior proportions:
  - 10 out of 30 fairy folk with crowns are princesses

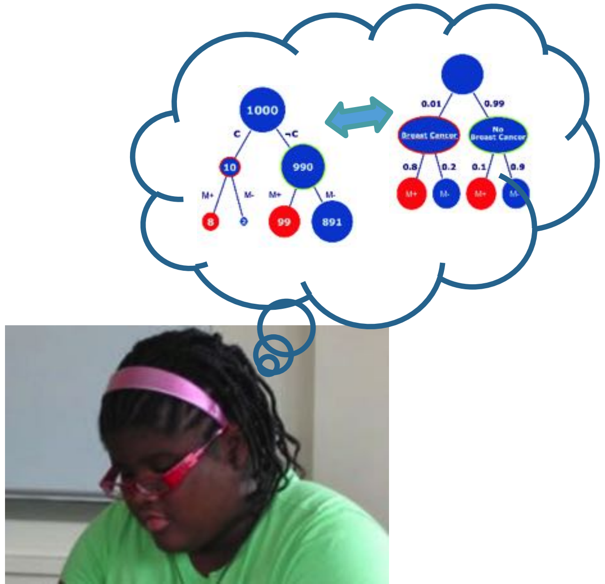
- Prior probability:
  - P(Princess) = 0.20
- Likelihood:
  - P(Crown | Princess) = 0.5
  - P(Crown | Mermaid) = 0.25
- Bayes Rule:
  - $$P(\text{Princess} \mid \text{Crown}) = \frac{P(\text{Crown} \mid \text{Princess})P(\text{Princess})}{P(\text{Crown})} = \frac{0.5 \times 0.2}{0.3} = 0.33$$

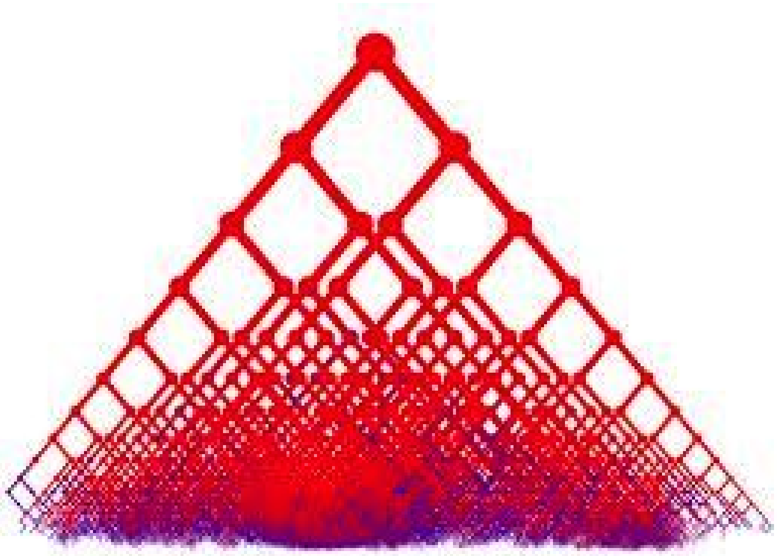
$$P(\text{Princess} \mid \text{Crown}) = \frac{P(\text{Crown} \mid \text{Princess})P(\text{Princess})}{P(\text{Crown} \mid \text{Princess})P(\text{Princess}) + P(\text{Crown} \mid \text{Mermaid})P(\text{Mermaid})}$$

$$= \frac{0.5 \times 0.2}{0.5 \times 0.2 + 0.25 \times (1 - 0.2)} = 0.33$$

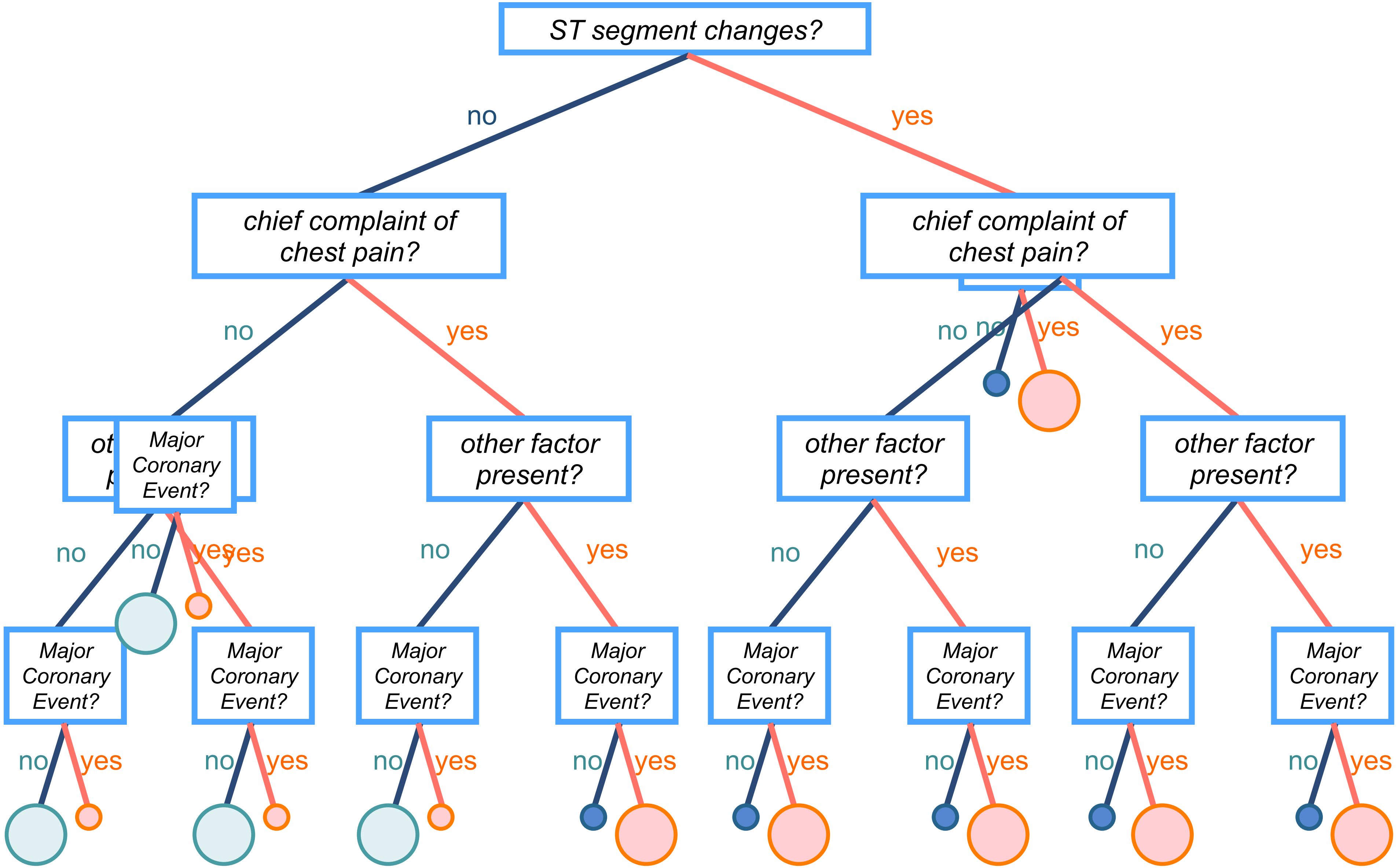
Natural Frequency to Probability

- Enactive and iconic representations with natural frequencies help young children develop intuitions about probability
- Older children can transfer understanding of natural frequency trees to probability trees
- But frequency trees become intractable (even for computers) as the number of factors grows
- To reason with multiple factors, we need new representations



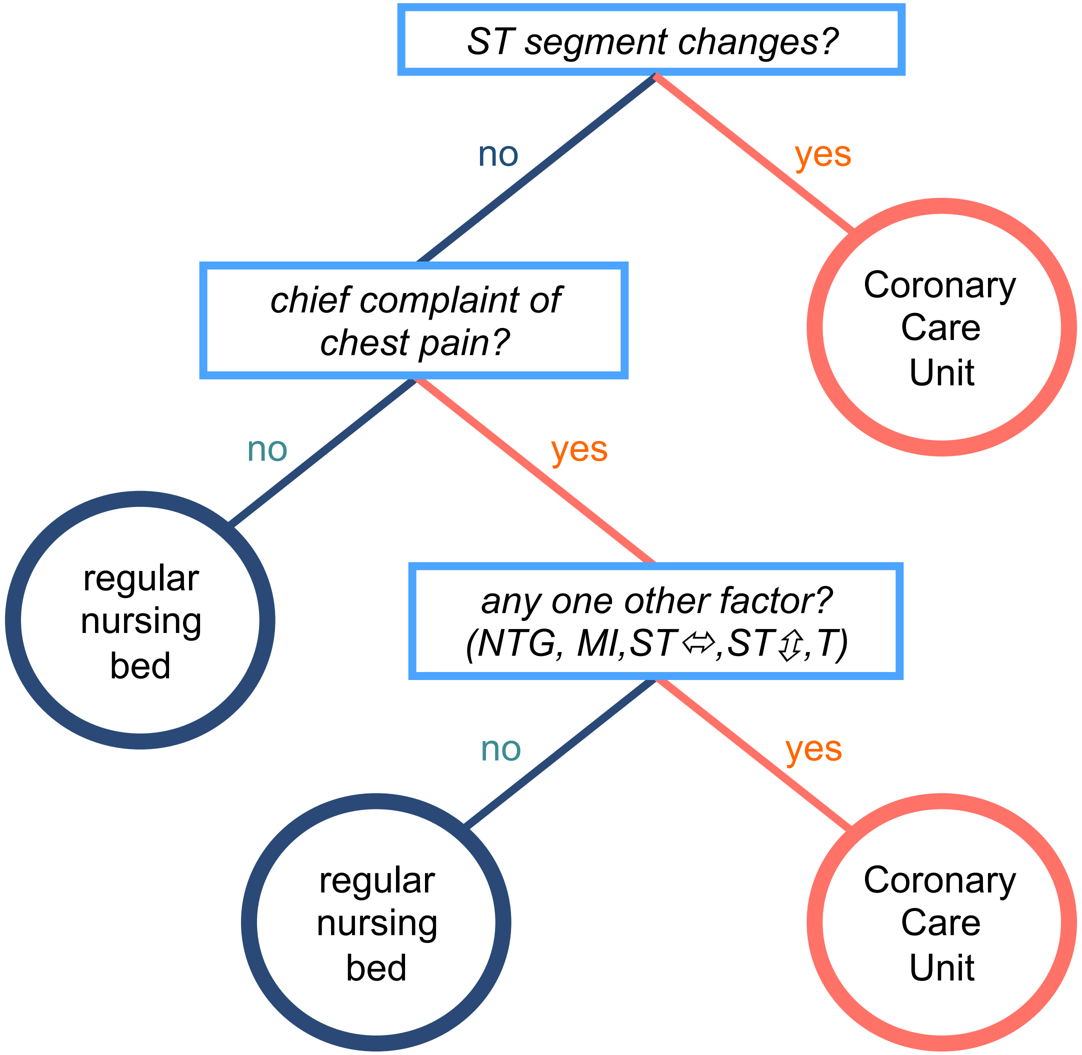


Diagnostic Frequency Tree → FFT



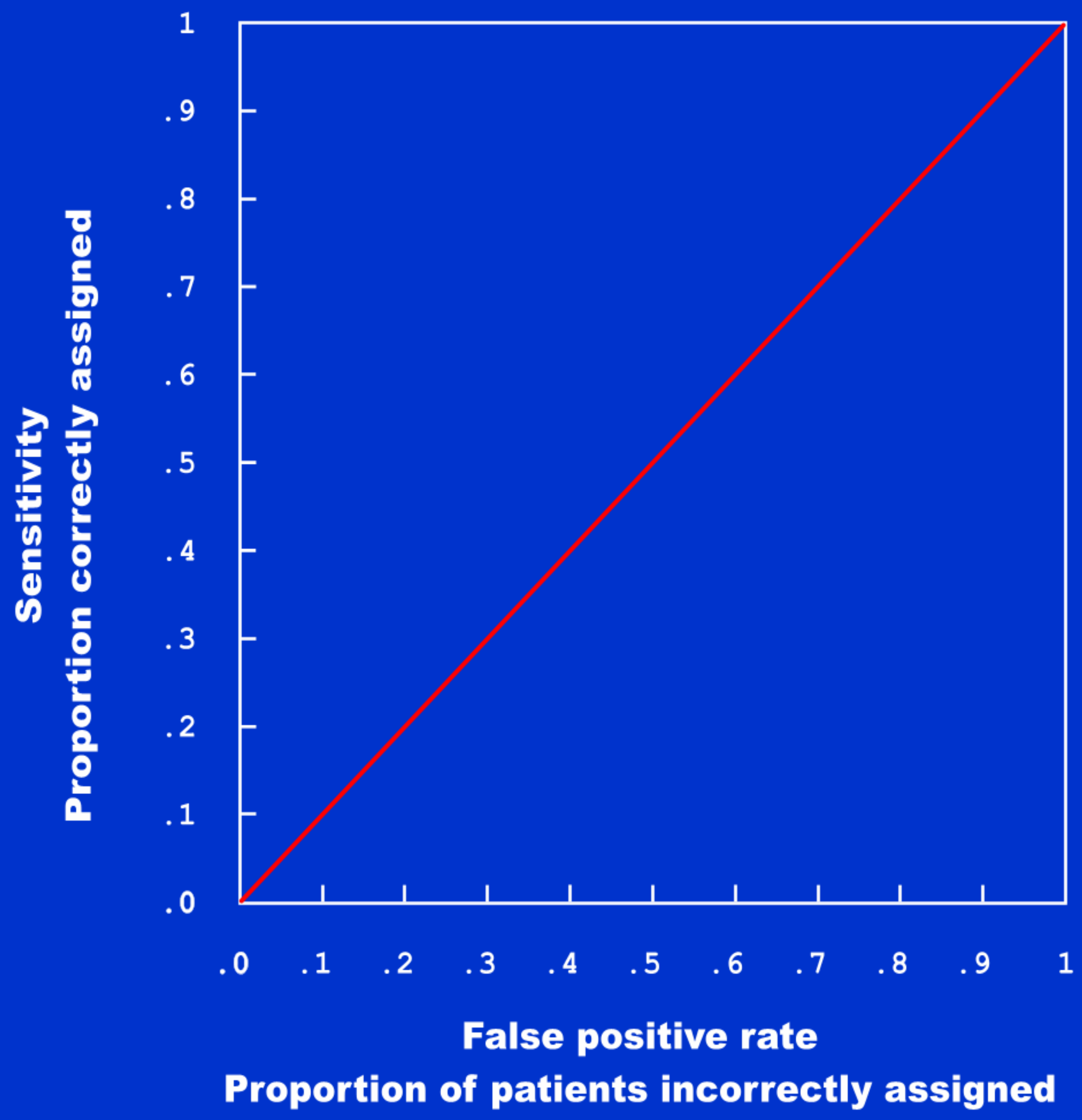
Fast and Frugal Tree

(Martignon, et al., 2003)

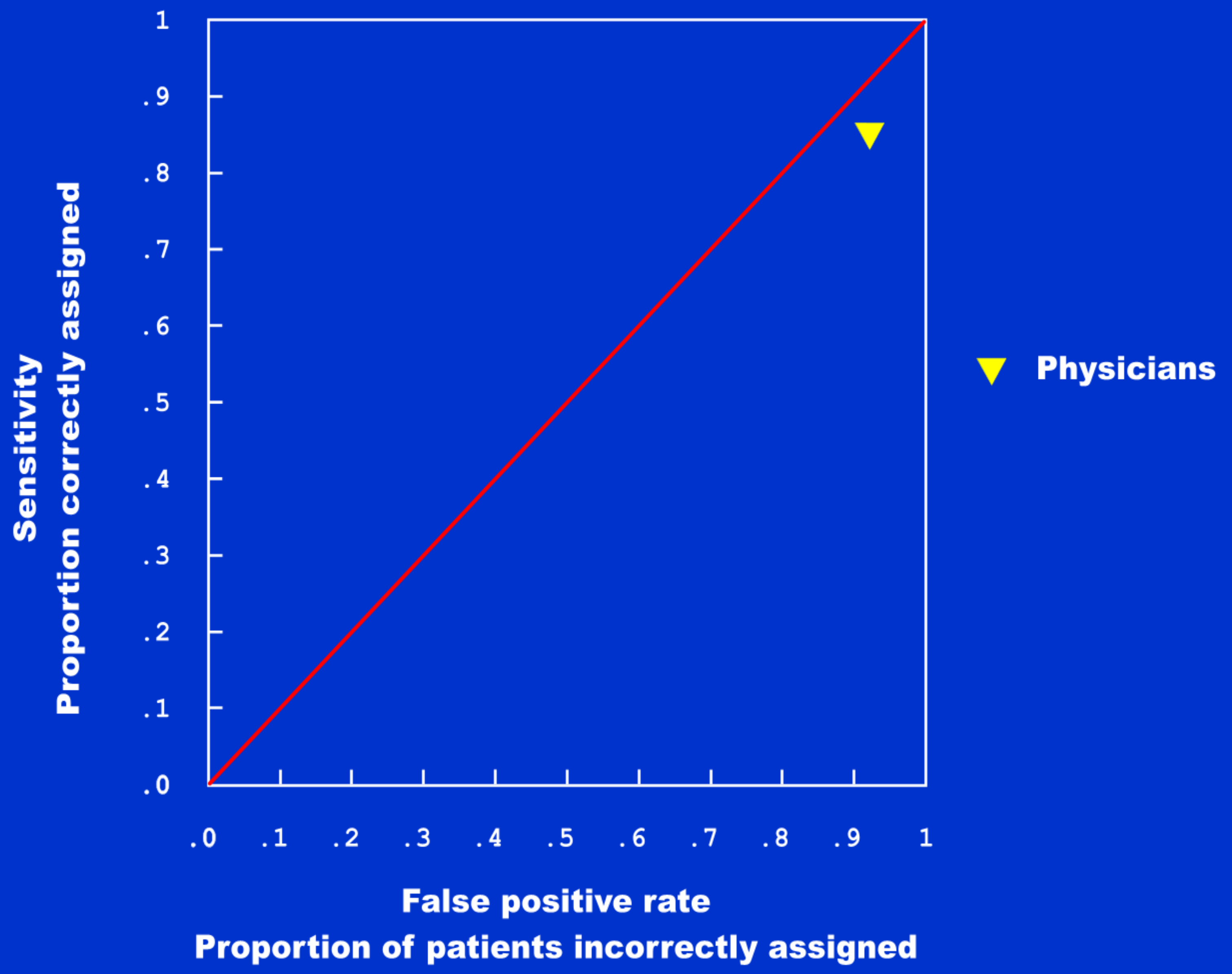


Example taken from Green & Mehr (1997)

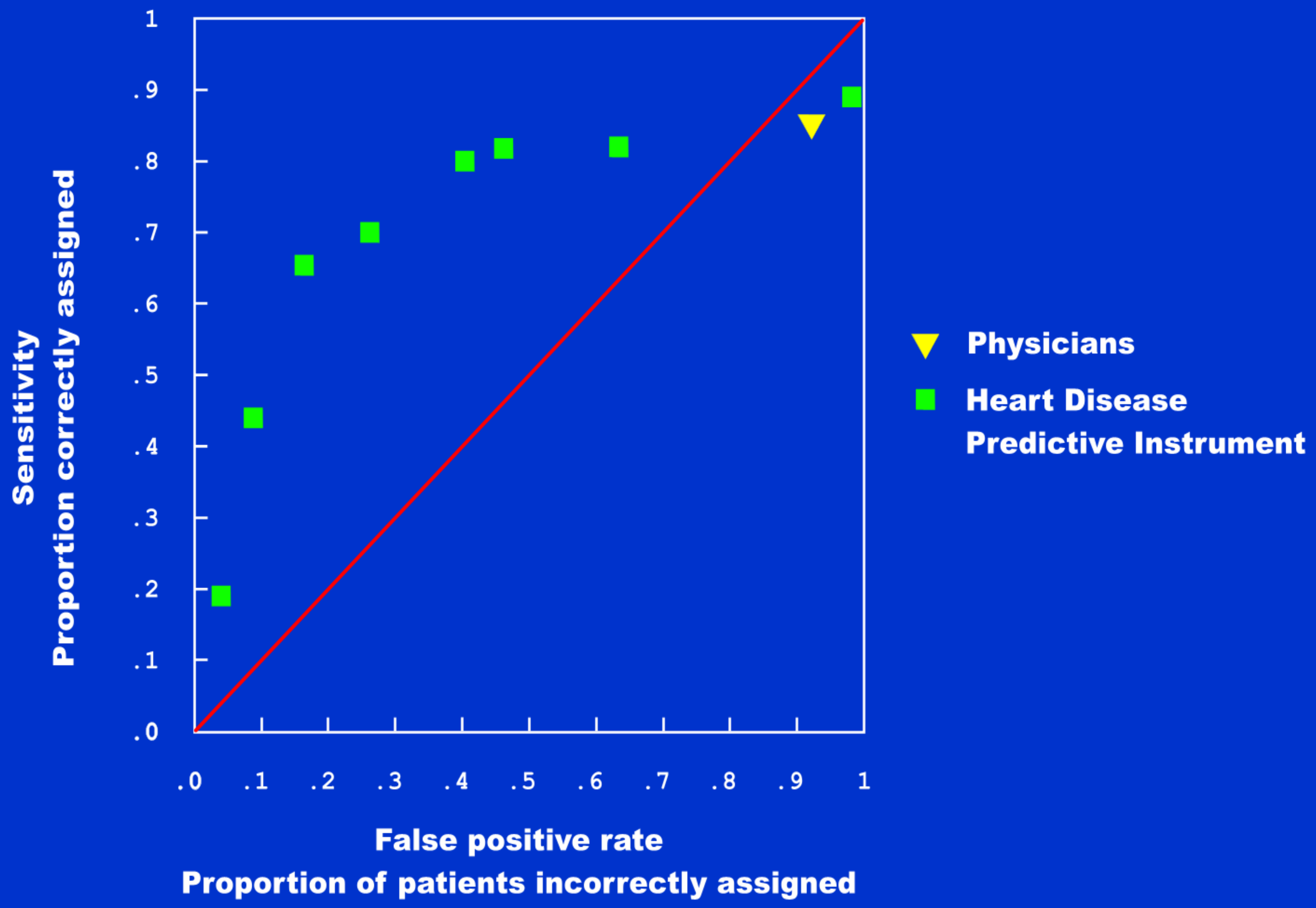
Emergency Room Decisions: Admit to the Coronary Care Unit?



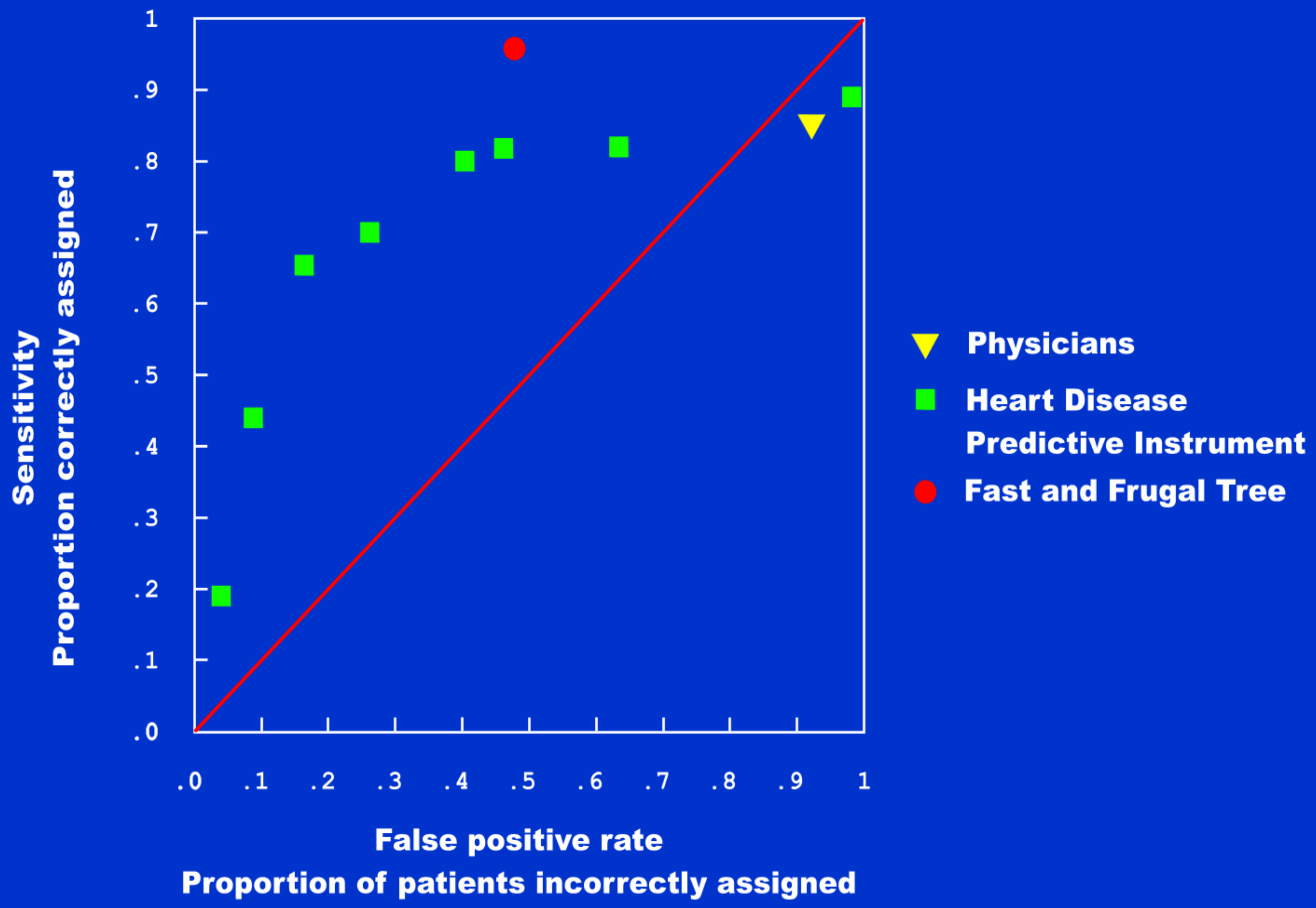
Emergency Room Decisions: Admit to the Coronary Care Unit?



Emergency Room Decisions: Admit to the Coronary Care Unit?

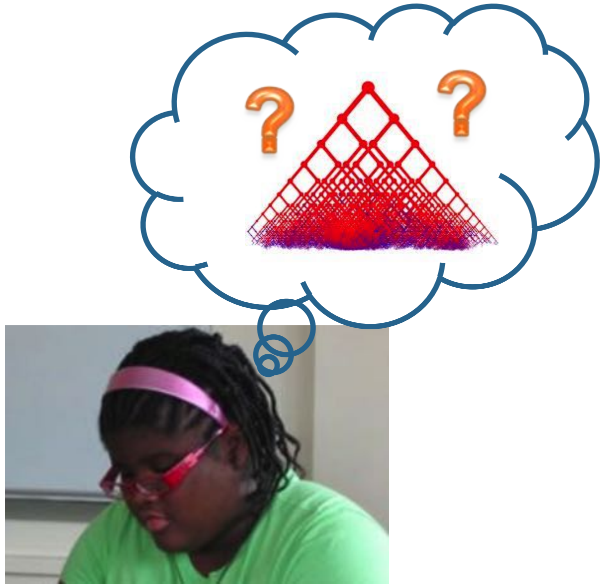


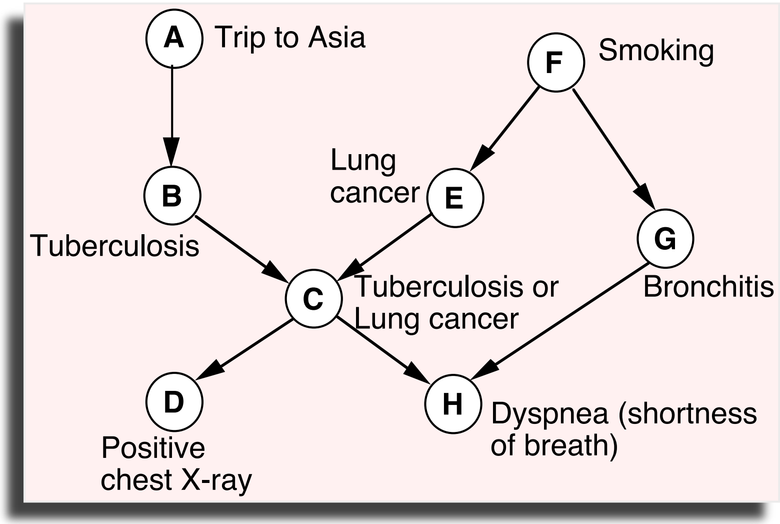
Emergency Room Decisions: Admit to the Coronary Care Unit?



Bayesian Reasoning with Many Factors

- Graphical models (Bayesian networks) use conditional independence to simplify specification and inference for probability models with many variables
- Can this powerful idea be exploited to teach more complex probability problems in school?





from Lauritzen and Spiegelhalter, 1988

Naïve Bayes

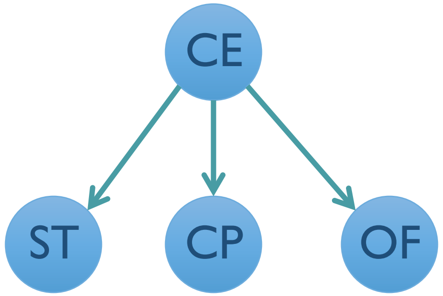
- Simple and robust Bayesian network model
- Assumes sensitivity and specificity of each factor do not depend on values of other factors, e.g.:
  - P(CP | CE, ST=yes) = P(CP | CE, ST=no)

CE = Coronary event

ST = ST segment elevated

CP = Chest pain primary symptom

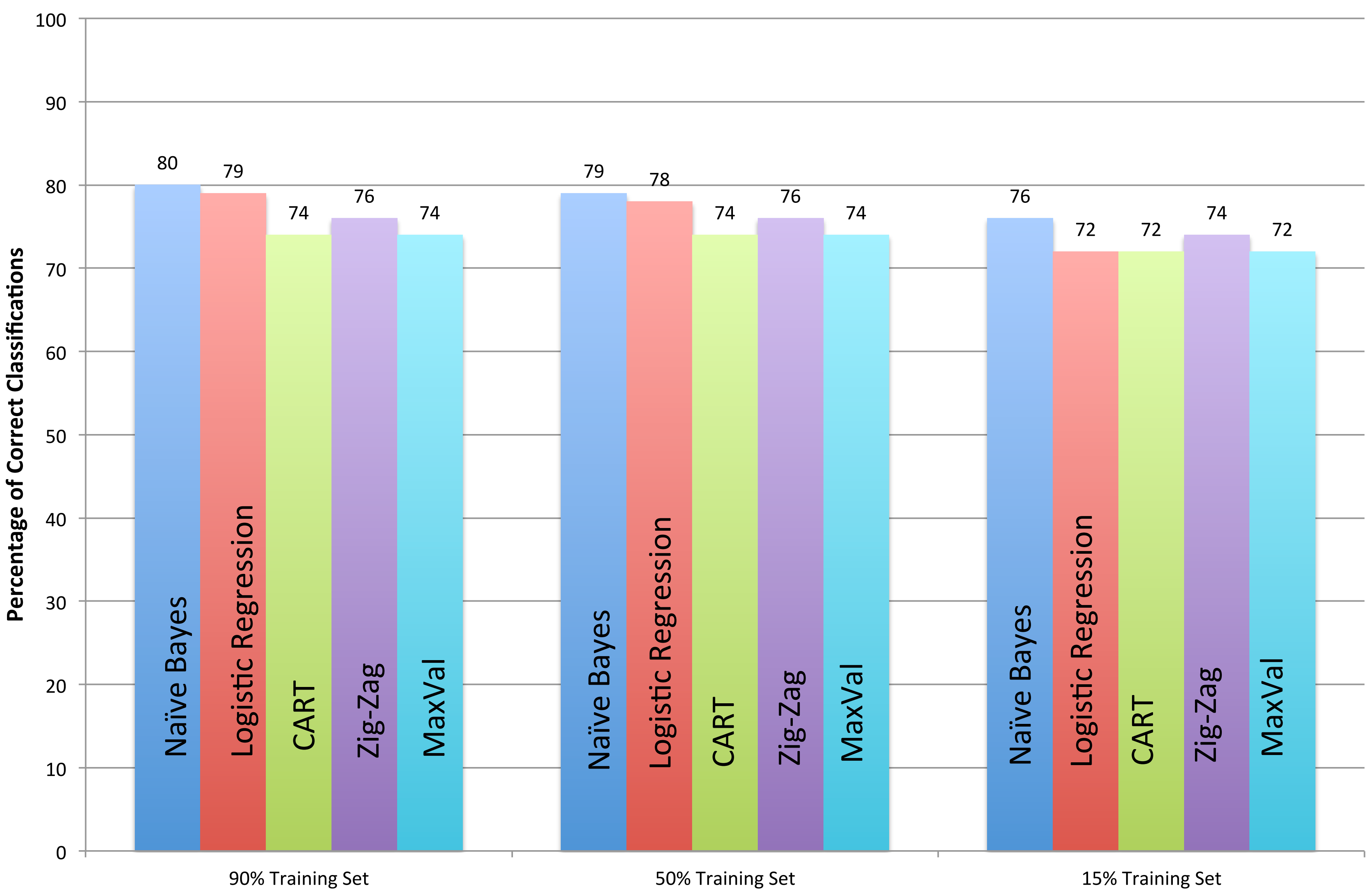
OF = Other factor present



- Use Bayes rule to find probability of CE given ST, CP and OF

Comparative Study

- 30 medical data sets
    - Most from UCI repository
    - Sample sizes ranged from 62 to 768
    - All had binary criterion
    - Continuous features were converted to binary features using median as boundary
- Five inference methods
    - F&FT with Zig-Zag
    - F&FT with MaxVal
    - CART
    - Logistic regression - most likely outcome was selected
    - Naïve Bayes - most likely outcome
- Each method was trained on a subset of the data and tested on remaining data
- This process was repeated 1000 times


Percentage of Correct Classifications
100
90
80
70
60
50
40
30
20
10
0
80
79
74
76
74
79
78
74
76
74
76
72
72
74
72
Naïve Bayes
Logistic Regression
CART
Zig-Zag
MaxVal
Naïve Bayes
Logistic Regression
CART
Zig-Zag
MaxVal
Naïve Bayes
Logistic Regression
CART
Zig-Zag
MaxVal
90% Training Set
50% Training Set
15% Training Set

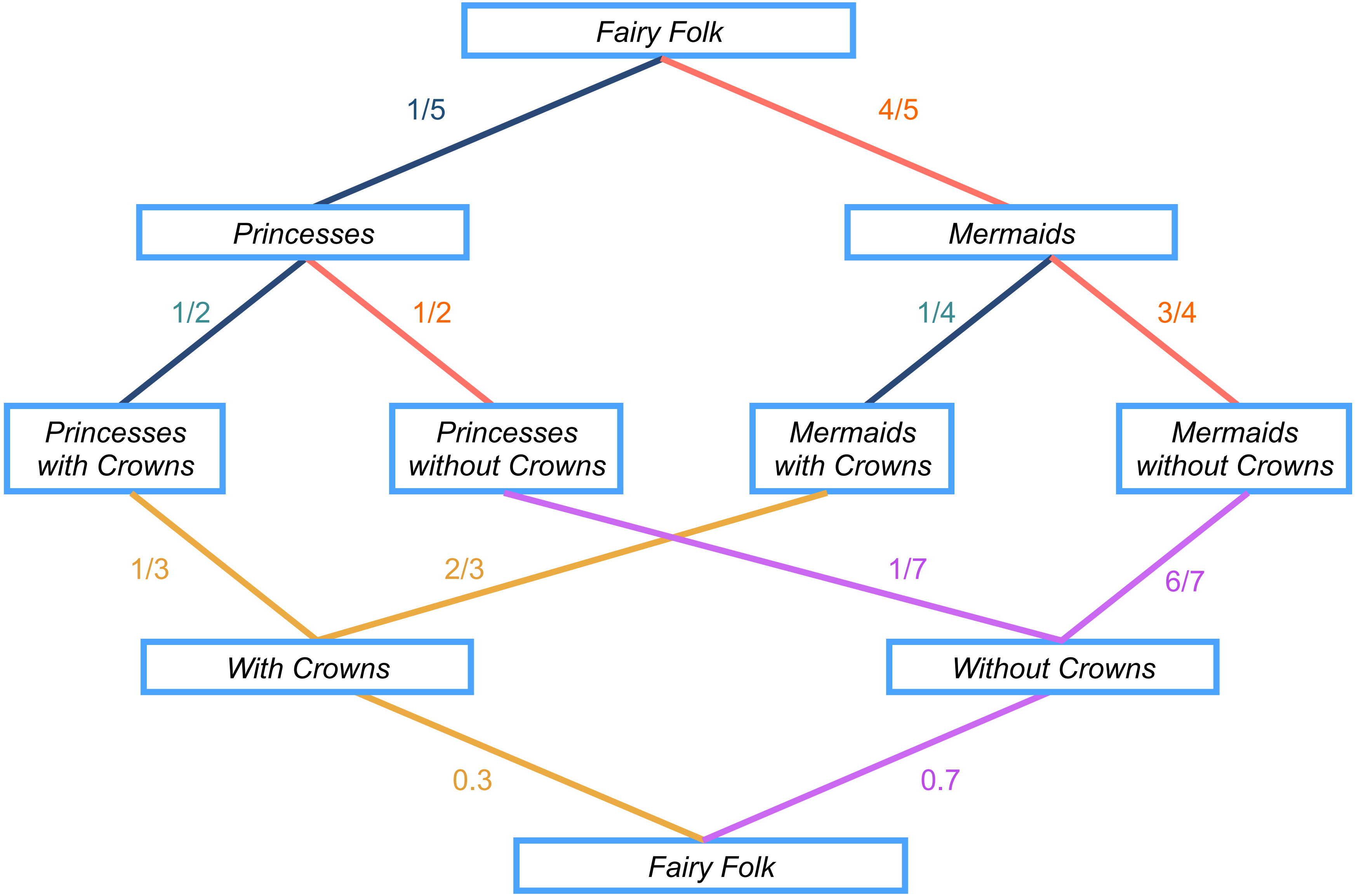
Remarks

- Naïve Bayes performed best overall
  (*Independence assumptions were clearly violated in all data sets*)
- All methods -- including F&F trees -- had comparable performance but FFT are much simpler and natural than all other algorithms
- As psychological studies have shown FFT's are akin to the human mind and mimic precisely what people tend to do

Conclusion

- Enactive and iconic representations with natural frequencies have been shown to foster intuitions about proportions (Martignon and Krauss, 2009)
- Training on natural frequencies and on transfer to probability fosters learning of probability (Wassner, Biehler, and Martignon, 2002)
- Fast and frugal trees are commonly used to perform diagnostic reasoning in medical and other domains (Fischer, et al., 2003)
- Naïve Bayes is a simple but powerful Bayesian model that (slightly) out-performs F&F trees on the study reported here
- Students trained with natural frequencies and F&F trees have the ingredients needed to understand and apply Naïve Bayes
- Studies are needed to evaluate instructional strategies for teaching Naïve Bayes in school

Diagnosis: Inverting the Tree



Diagnosis: Inverting the Tree