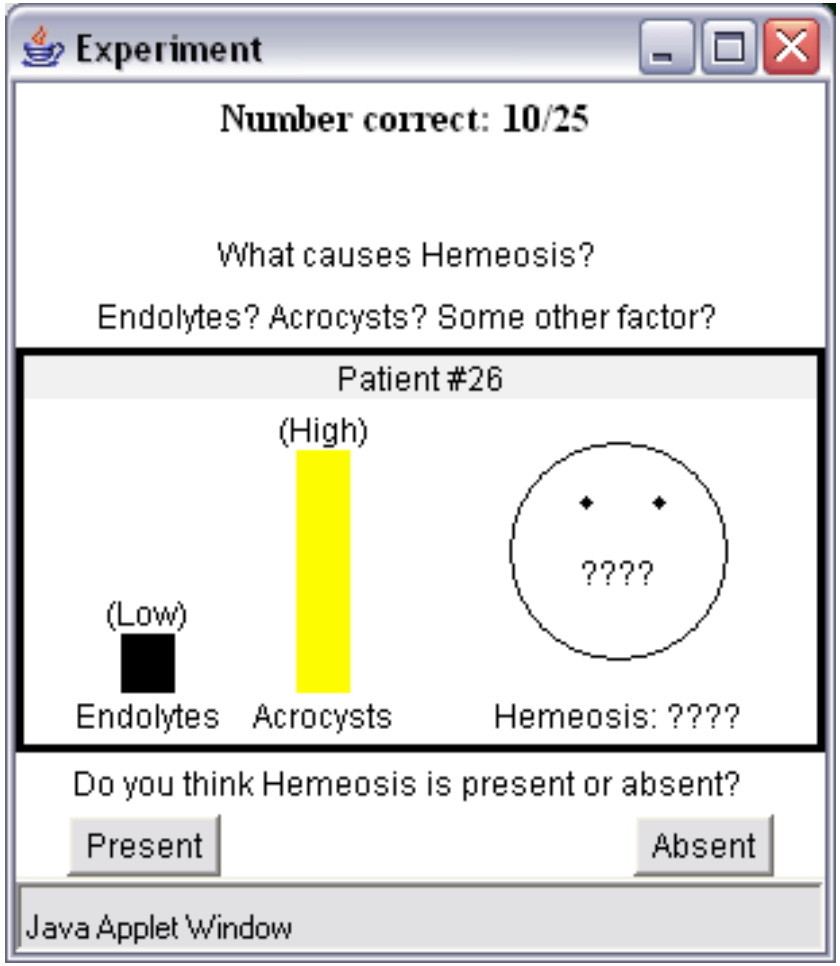
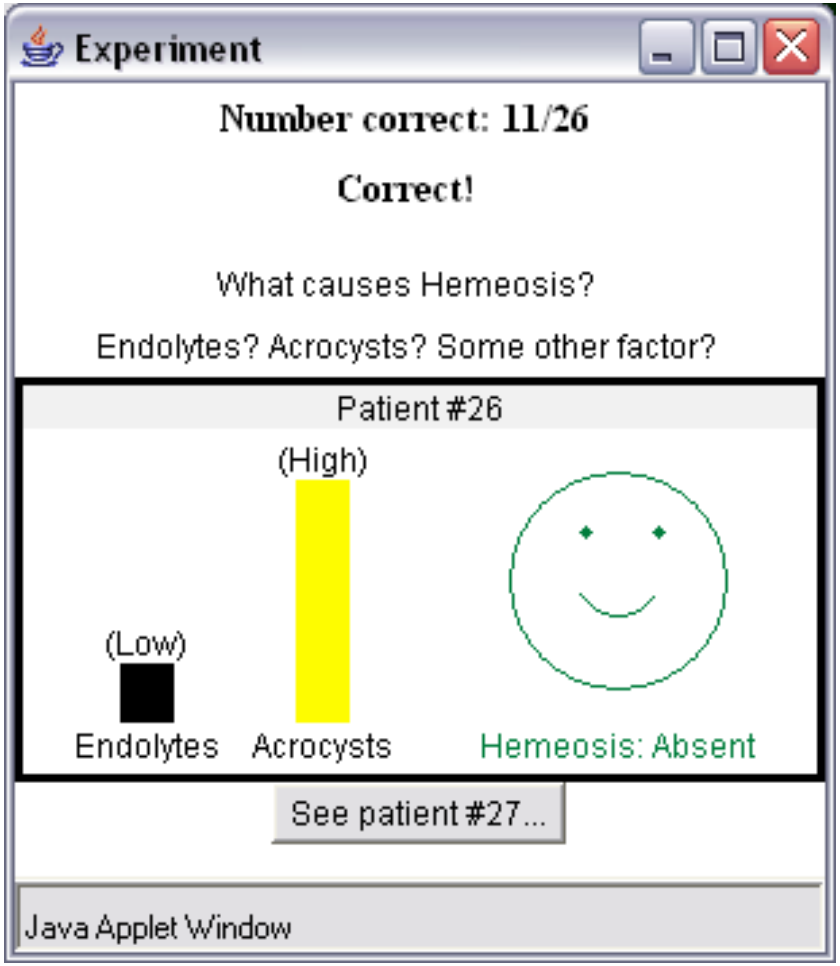
Psychological Models of Causal Inference

Paradigmatic Experiment

- A set of binary potential causes: C_1, …, C_n
- A known binary effect: E
  - This distinction is given to the participant
  - Cover stories are explicitly designed to minimize the role of prior beliefs
- Observational data about variable values
  - Different formats: online, list, or summary

Paradigmatic Experiment





Paradigmatic Experiment

- Obtain verbal estimates of causal influence
  - Often, ratings on a -100 to +100 scale
  - Responses are somewhat sensitive to question
  - Occasionally, structural beliefs are obtained using a range of interfaces
- Goal of theories: model the (mean) ratings as a function of the observed series of cases (usually, just the probability distribution)
  - Or a series of (mean) ratings

Conditional ΔP Theory

- Predictions of causal strength judgments are only made in the large-sample limit, and are proportional to the conditional contrast(s):

$$\Delta P_{C.\{X\}} = P\,(E\,|C\ \&\ X\,) - P\,(E\,|{\sim}C\ \&\ X\,)$$

- This theory makes no determinate prediction if no conditional contrasts are defined, or if the contrasts for a potential cause vary

Causal Power Theory

- Events are due to unobservable causal capacities (*a la* Cartwright)
- Causal learning is the estimation of the frequency of operation of these capacities

Generative Cause

$$p_C = \frac{\Delta P_C}{1 - P(E \mid \neg C)}$$

Preventive Cause

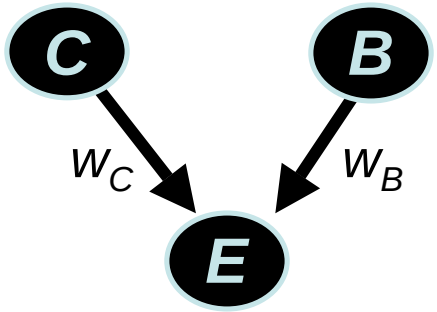
$$p_C = \frac{-\Delta P_C}{P(E \mid \neg C)}$$

Central Claim

- Bayes nets provide a good model of people's causal beliefs
  - All of the non-Bayes net psychological theories are parameter estimators for a particular, fixed-structure Bayes net
  - There are interesting psychological theories of causal learning based on Bayes nets

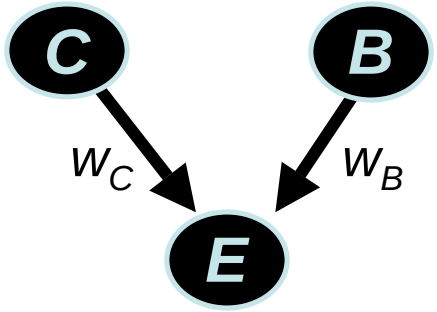
Unifying the Other Theories

- Claim: Causal Bayes nets can provide a framework to unify the other theories.



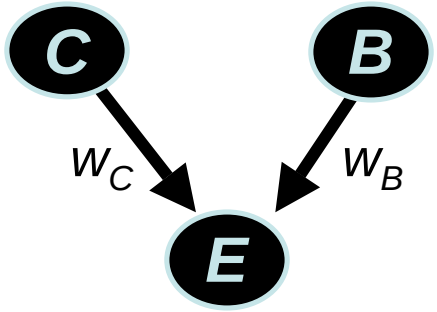
- Assume that *B* always occurs
- Specify $P(E) = f$ (*C*'s occurrence, w_C, w_B)
  - Not the only way to specify it, but convenient

Unifying the Other Theories



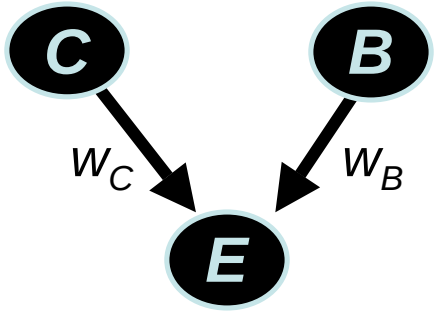
- Suppose $P(E) = w_B + w_C \times \delta(C)$
  - where $\delta(C) = 1$ if C occurs, 0 otherwise
- ΔP is the maximum likelihood estimate of w_C
  - So ΔP can be reinterpreted as a parameter estimator for a fixed-structure, fixed-functional form Bayes net!

Unifying the Other Theories



- Or: $P(E) = w_B + w_C \times \delta(C) - w_B \times w_C \times \delta(C)$
- Causal power is the maximum likelihood estimate for w_C in *this* equation
  - I.e., Power PC is also a parameter estimator

Unifying the Other Theories



- ΔP and power PC (and their dynamical versions) are doing exactly the same thing, except that they assume different underlying functional forms

Potential Problem

- Observing a variable and manipulating that variable are not the same
- The observation / manipulation distinction is essentially not drawn in the parameter estimation theories
  - So how could they possibly be right?

Interventions & Estimators

- Interventions only break incoming arrows, so observation and intervention are the same for *exogenous* variables
  - I.e., variables with no incoming arrows
- We can draw exactly the same inferences from (a) observation of an exogenous variable; and (b) intervention on one

Intervention & Estimators

- It’s fine for the psychological theories to ignore the observation / manipulation distinction, because:

  There is no observation / manipulation distinction in the assumed, fixed-structure causal Bayes net!

Bayes Net Psychology

- Two different strategies:
  - Rational analysis of human causal learning
    - Model people as using a causal Bayes net structure learning algorithm, even though they might actually be using heuristic methods
  - Causal model theory
    - People start with an initial causal structure, try to fit the observed data to it, and change their minds only if the data are sufficiently anomalous

Bayes Net Psychology

- Two types of experimental tests
  - Prediction from - and use of - interventions
  - Learning integrated causal structures
- Data from both types of experiments support richer Bayes net theories (and not just parameter estimation)