

Putting the Pieces Together: Case Studies

Case Studies

- College student drop-out
- *Spartina* grass growth
- Climatological structure
- Stress, depression, and religion
- Test-taking anxiety

College Student Drop-out

- Data from 1993-94 *U.S. News and World Report* surveys
 - Highly sub-optimal data for this problem
- PC algorithm (classical search) outputs:
Everything else — Average SAT → Dropout rate
- \Rightarrow Of the survey variables, Average SAT score is the only direct cause of Dropout

“Experimental” Test

- 1994: Carnegie Mellon University alters its financial aid policies to increase average SAT scores of freshman class
- After the policy change: Dropout rate since 1994 has decreased monotonically with increasing average SAT of freshman class

Spartina Grass Growth

- What factors influence the biomass of *Spartina* grass in the Cape Fear estuary?
 - In other words, what causes some *Spartina* plants to be larger, heavier, than others?



Many Possible Answers

- 13 chemical factors (biologist-identified):
 - Acidity (due to acid rain)
 - Salinity (due to proximity to the ocean)
 - Ammonia salts (due to agricultural drainage)
 - Phosphates
 - Nitrogen
 - Metal ions of many kinds
 - Etc.

Many Possible Answers

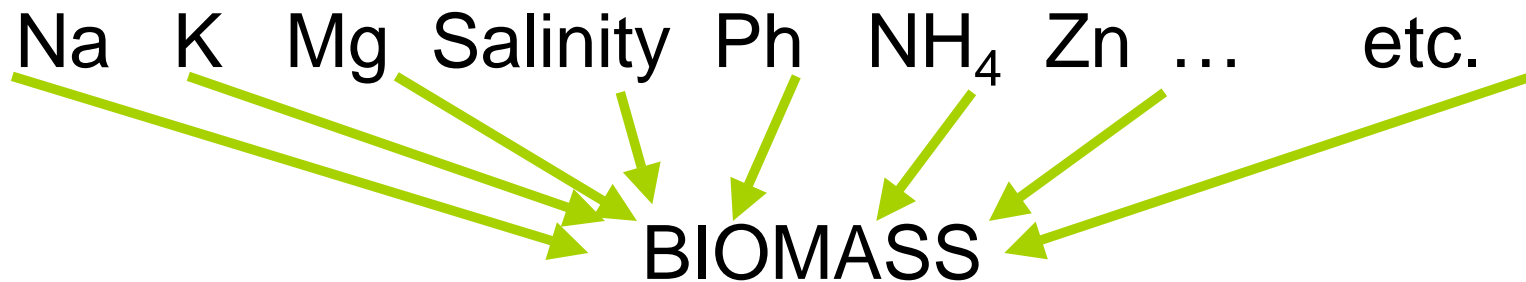
- Also plenty of non-chemical possibilities
 - Mechanical soil factors influencing aeration of the soil
 - Solar irradiation
 - Temperature
 - Etc.

Causal Learning Problem

- Determine the factors that influence *Spartina* biomass from measurements of only 45 plugs sampled from the estuary
 - Very small sample size given the number of possibly relevant variables

Causal Learning Problem

- Start with this potential causal model:



- And then find out:
 - Which causal arrows actually exist; and
 - Estimate how BIOMASS depends on the corresponding factor

Spartina Grass Growth

- Regression analyses give different answers depending on included variables and particular techniques
- Constraint-based search algorithm concluded that the only variable that directly influences Biomass is pH
 - Other variables indirectly influence Biomass, but always through the mediating factor of pH
 - Causal structure is just like the dropout case

Spartina Grass Growth

- Linthurst (the biologist who did the original study) was convinced that salinity has a direct effect on biomass, and so did a randomized greenhouse experiment
- Experiment results:
 - pH influences biomass when salinity is controlled for
 - Salinity has no influence on biomass if pH is held constant

Climatological Structure

- Structure of ocean teleconnections
 - 504 monthly measurements, 6 ocean indices
 - **QBO** (Quasi Biennial Oscillation); **SOI** (Southern Oscillation); **WP** (Western Pacific); **PDO** (Pacific Decadal Oscillation); **AO** (Arctic Oscillation); **NAO** (North Atlantic Oscillation)
 - Some *highly* non-linear relationships
 - Temporal order is the only prior knowledge
 - Ocean indices were defined by domain experts, but we were told nothing about their structure

Climatological Structure

- Learned structure:



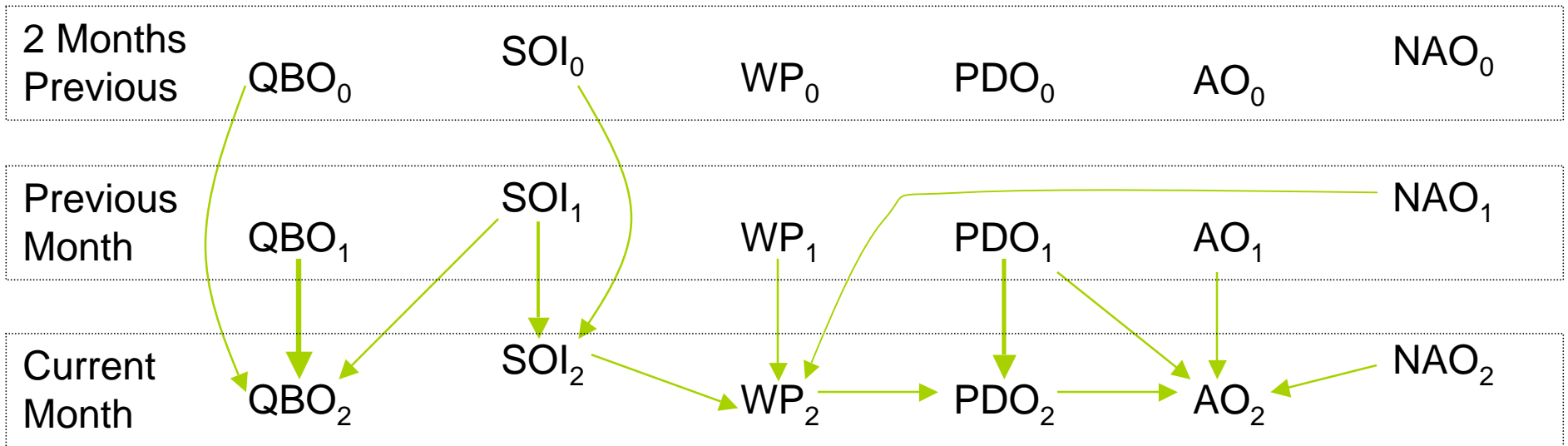
Time Direction

2 Months Previous	QBO_0	SOI_0	WP_0	PDO_0	AO_0	NAO_0
Previous Month	QBO_1	SOI_1	WP_1	PDO_1	AO_1	NAO_1
Current Month	QBO_2	SOI_2	WP_2	PDO_2	AO_2	NAO_2

Climatological Structure

- Learned structure:

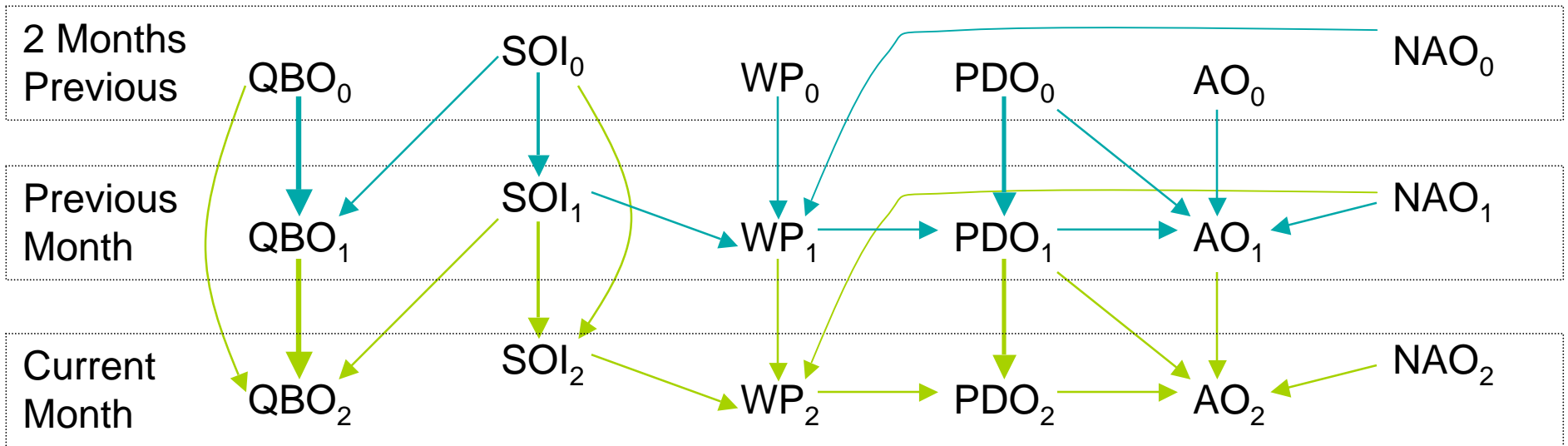
↓ Time Direction



Climatological Structure

- Learned structure:

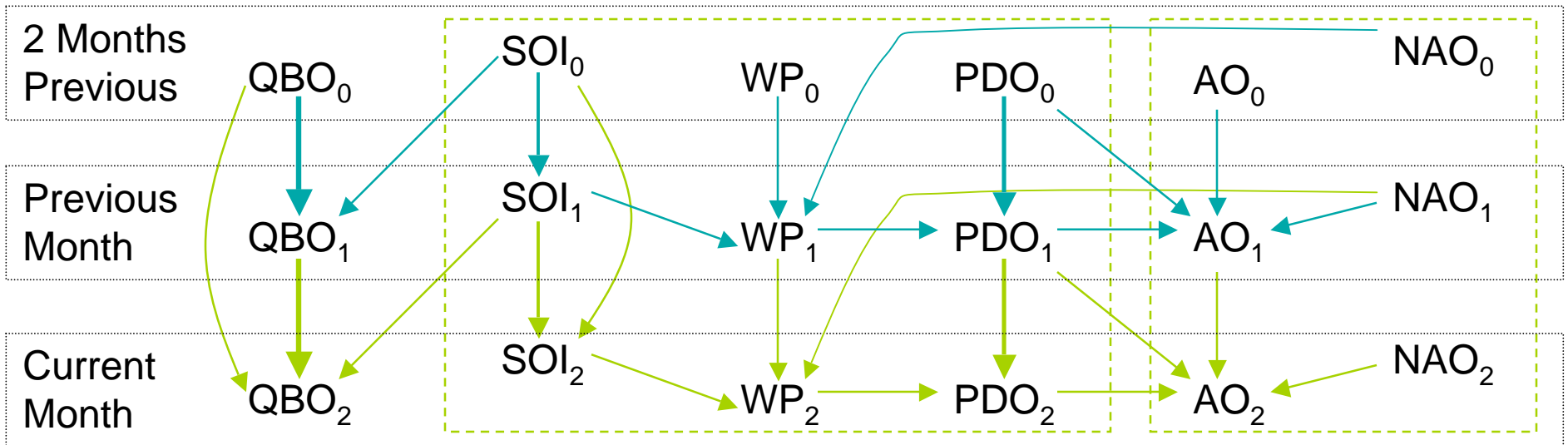
↓ Time Direction



Climatological Structure

- Learned structure:

↓ Time Direction



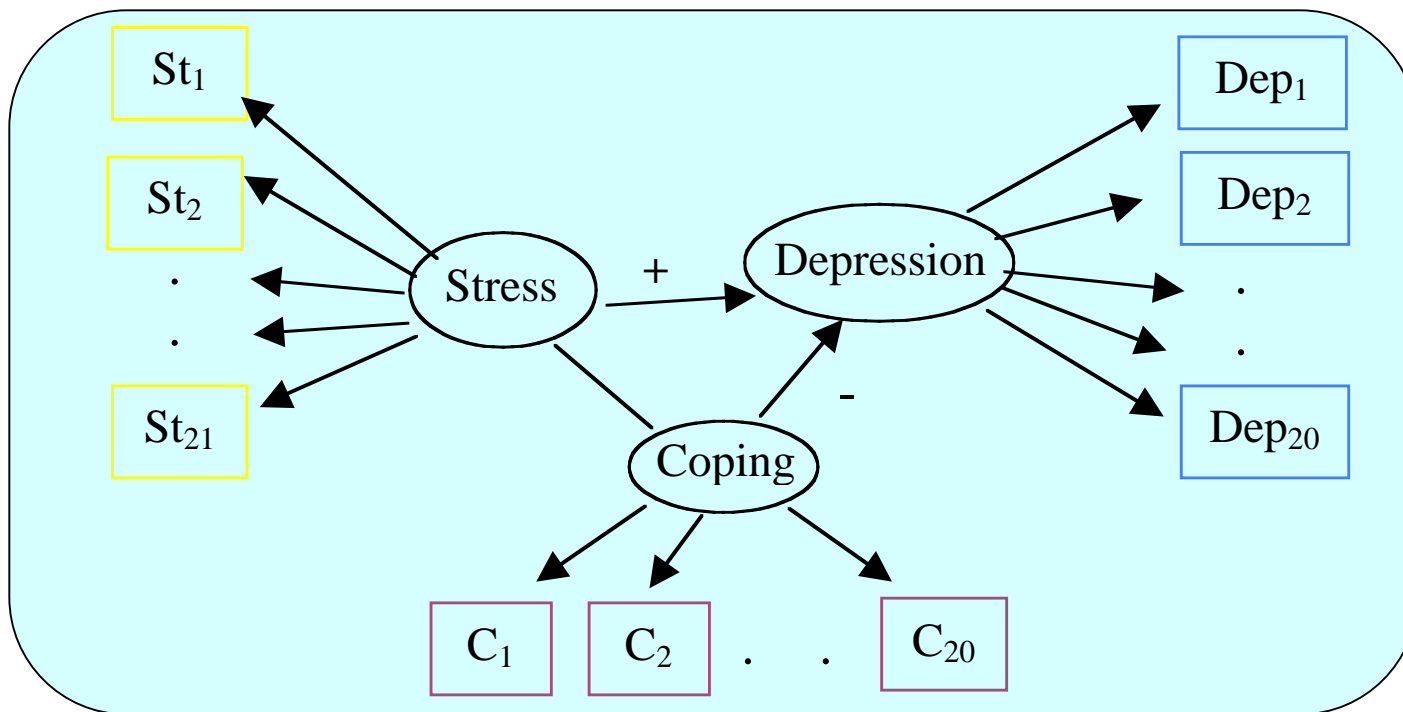
All causal relationships are either known or plausible (according to domain experts)

Stress, Depression, & Religion

- What is the connection between stress, depression, and religious coping?
 - We might think that religious people are better able to handle stress, and so less likely to suffer from depression
 - Is this actually correct?
 - 61-item survey experiment with 127 MSW students

Domain Model

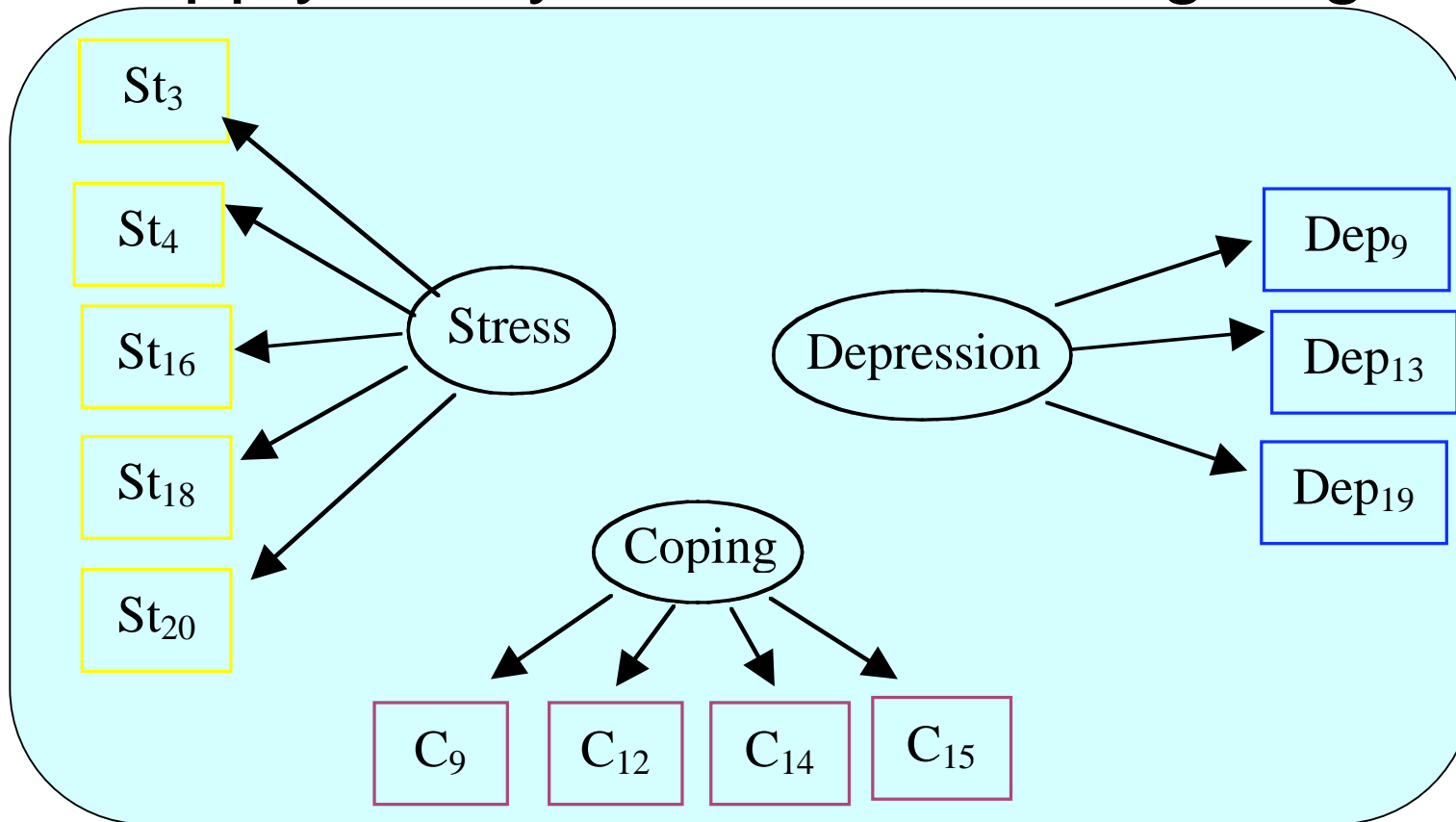
- Domain expert's model:



Statistical tests strongly reject this model

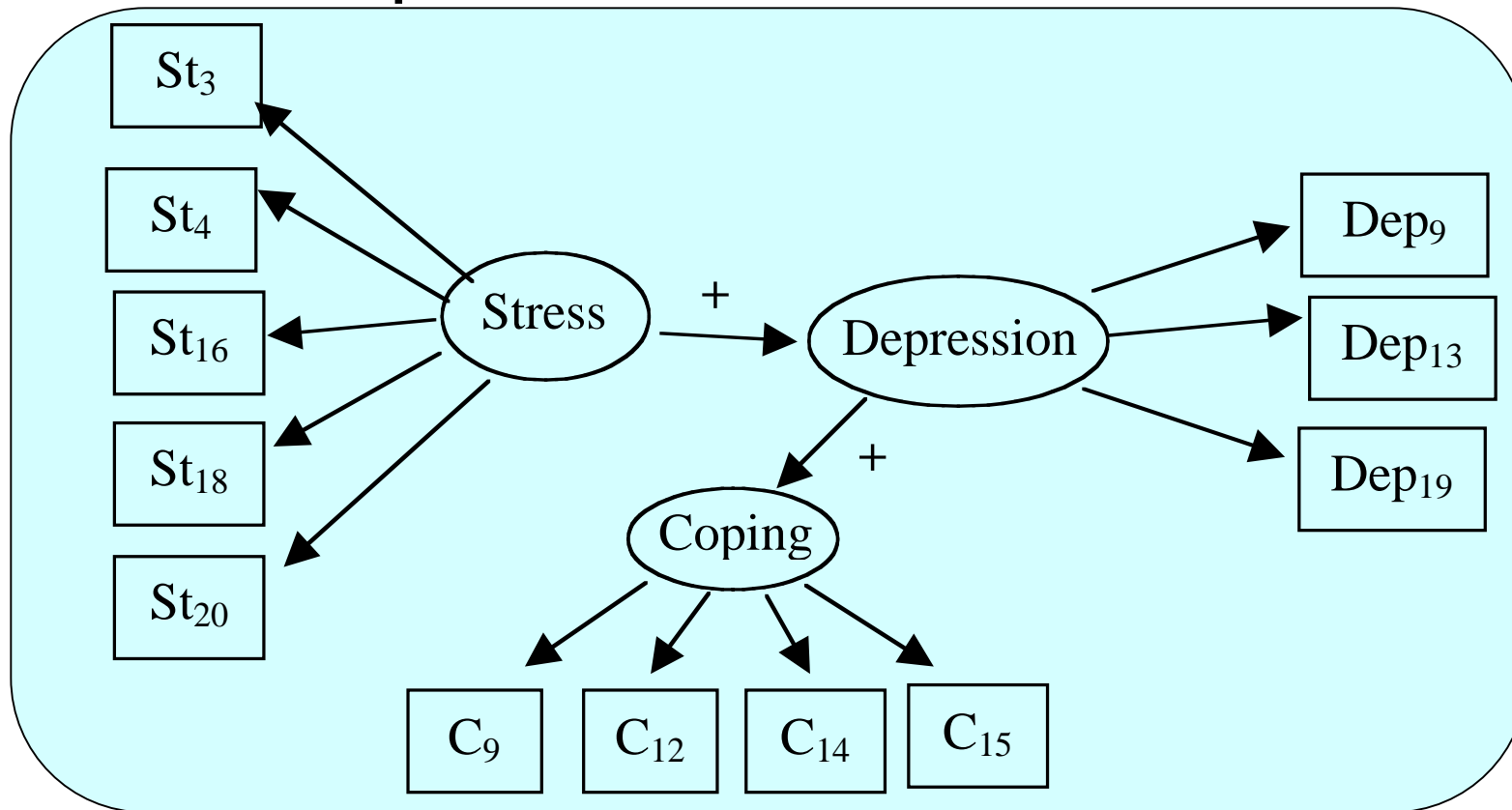
Latent Variable Discovery

- Apply a Bayes Net “clustering” algorithm:



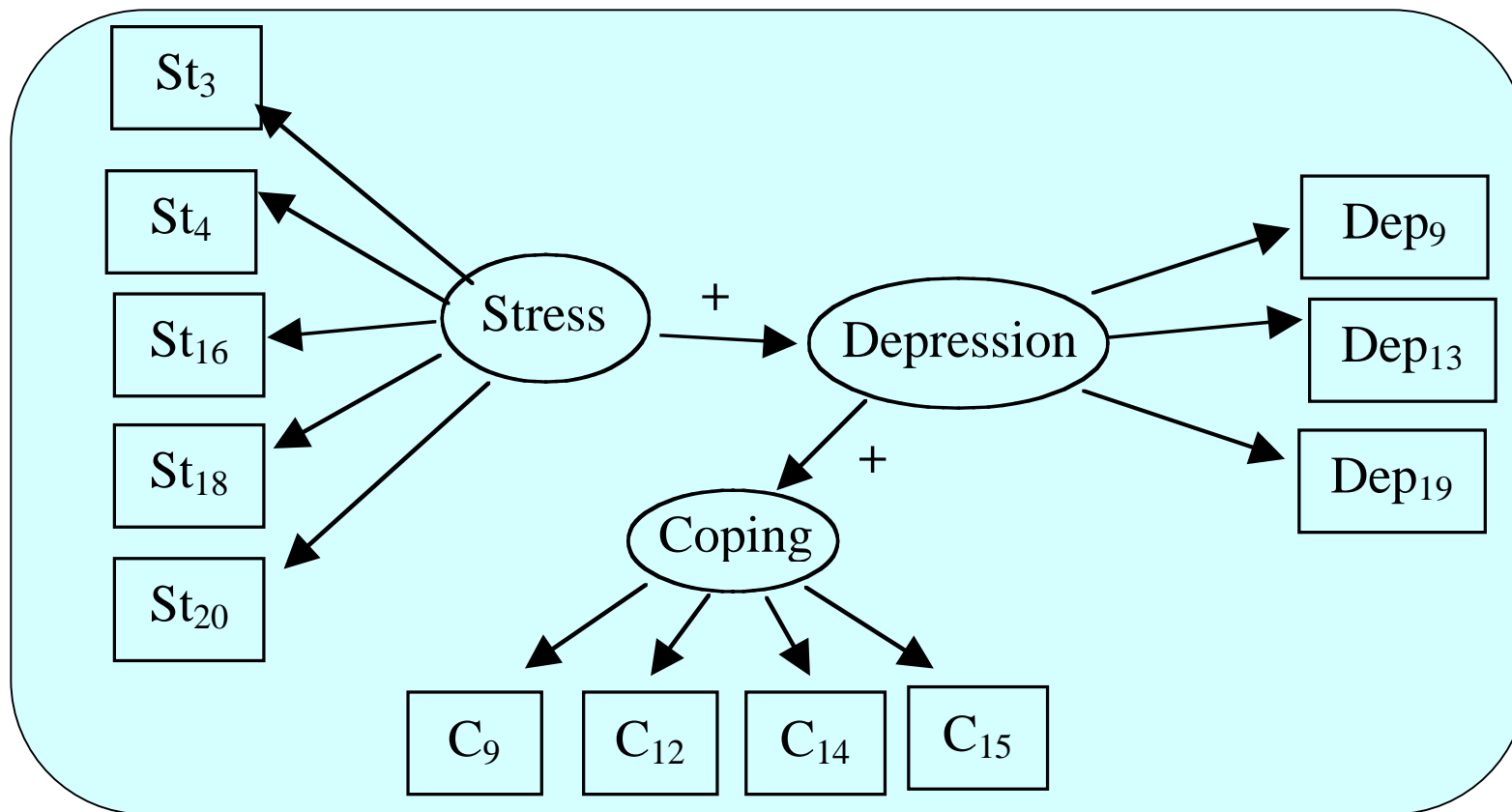
Latent Variable Structure

- If S is prior to D , then the structure is:



Latent Variable Structure

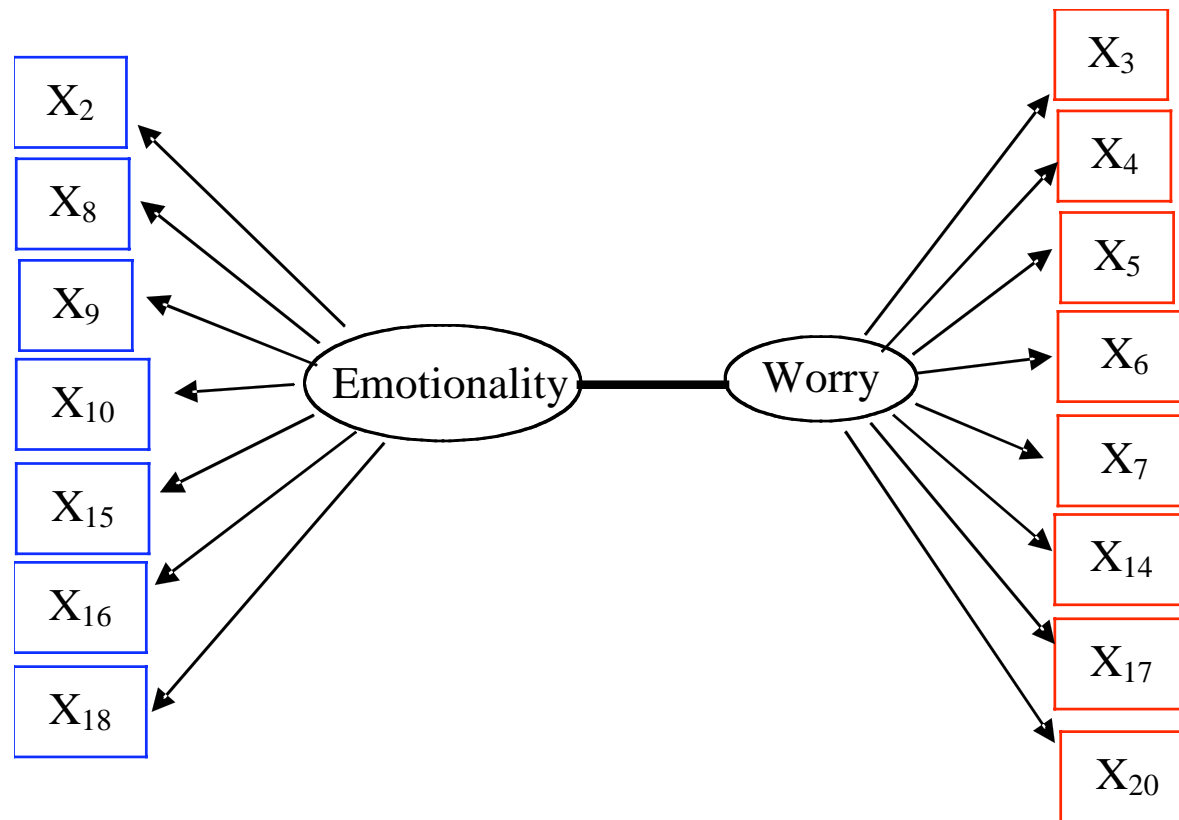
- ...and it cannot be statistically rejected



Test-Taking Anxiety

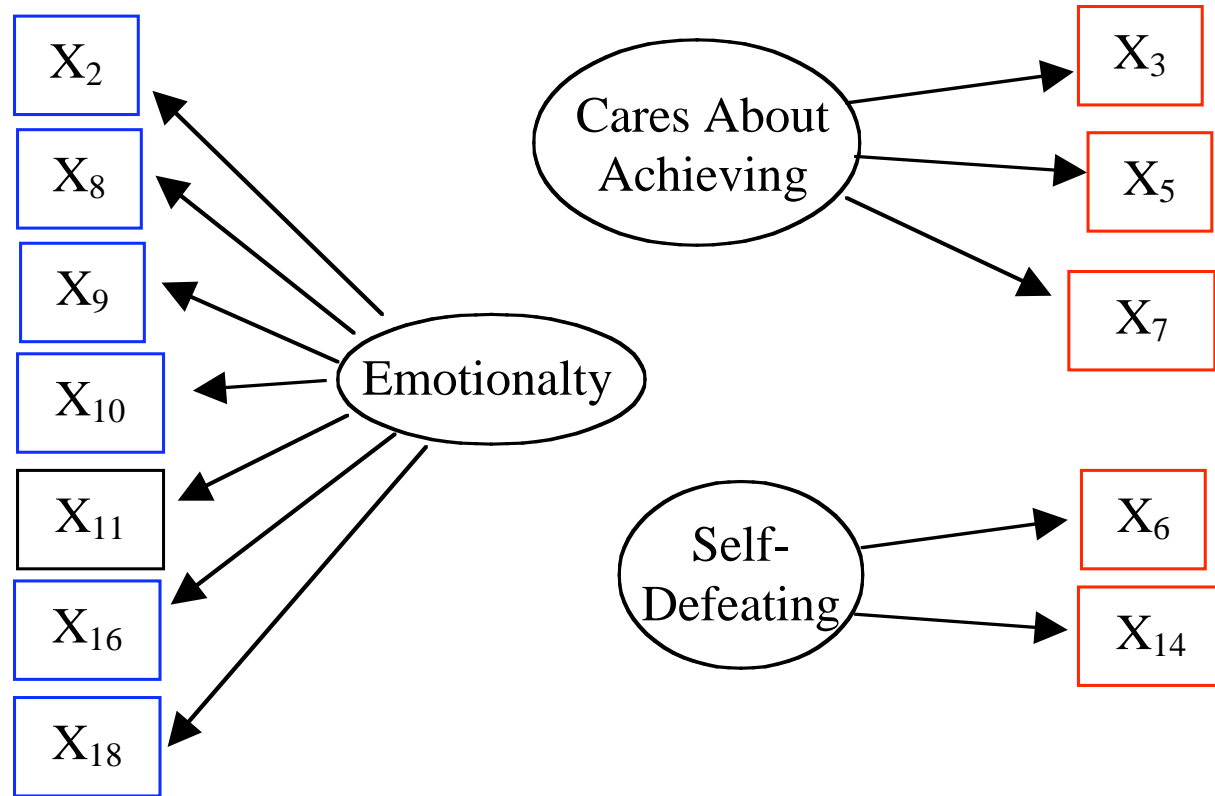
- Why do some people suffer from severe anxiety when they have to take a test?
 - Is it due to “local” factors?
 - Is it due to persistent features of personality?
 - 20-item survey experiment with 335 high school students (British Columbia)

Model from Factor Analysis



- Latent variable names are given by domain scientists
- Factor Analysis model is highly significantly *rejected*

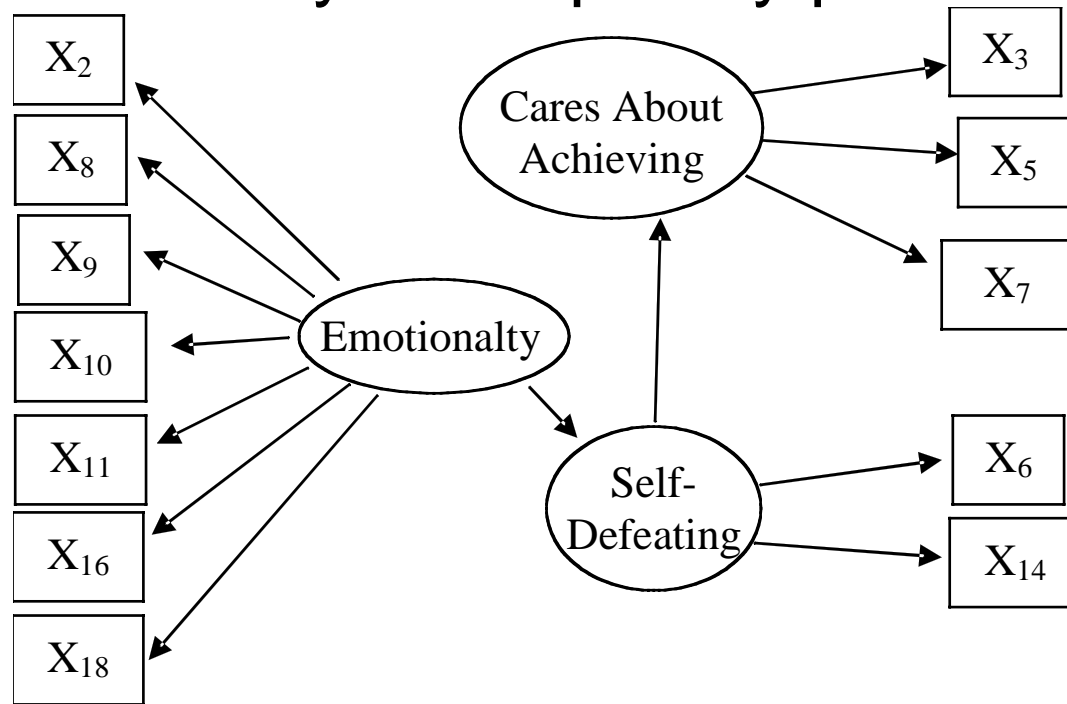
Bayes Net Clustering Model



- The latent variable names are ours (based on the survey items that fall into each cluster)

Latent Variable Structure

- If Emotionality is temporally prior:



- This model cannot be statistically rejected!