CNBC CENTER FOR THE NEURAL BASIS OF COGNITION
Carnegie Mellon University

Individual Variation in Causal Learning

Laila Johnston[1,3], Noah Hillman[2,3], David Danks[3,4,5]
[1] Department of Mathematics, UCF; [2] Department of Mathematics, Statistics, and Computer Science, St. Olaf College;
[3] Center for the Neural Basis of Cognition, CMU; [4] Department of Philosophy, CMU; [5] Department of Psychology, CMU
UCF UNIVERSITY OF CENTRAL FLORIDA
ST. OLAF COLLEGE

Introduction

- Numerous causal learning models have been proposed to explain how individuals make inferences about the relationship between a cue and a subsequent effect
- **Goal:** Determine which causal learning models best predict subjects' responses while examining model consistency within subject across conditions

Data

Two experimental datasets from Danks & Schwartz (2005, 2006) were analyzed:

- Participants were presented with a sequence of binary cause-effect cases
- After each case, the participant estimates the strength of the relationship between the cause and effect on a scale from -100 to 100
- Sequences are non-stationary: at the halfway point in each trial, the causal relationship switches direction

Procedure

Generate Predictions: A grid search was employed to generate predictions for reasonable sets of parameter values for each model

Calculate SSE: For each participant in each condition, the sum of squared errors between the participant data and the model predictions was found

Find Best Fit Model: The causal learning model with the set of parameters that generated the lowest SSE is declared the model of best fit for that participant in that condition

Causal Learning Models

Existing Models

Augmented Rescorla-Wagner
$$\Delta V_i = \alpha_{i\gamma(C_i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\sum_{\delta(C_j)} V_j)$$

Proportion of Confirming Instances (PCI)
$$\Delta V_j^{i+1} = \beta((-1)^{|\delta(E)-\delta(C_j)|}\lambda - V_j^i)$$

Catena Belief Adjustment
$$J_i = J_{i-1} + \gamma(PCI - J_{i-1})$$

Causal Support
$$\text{support} = \log\left(\frac{P(D|\text{Graph }1)}{P(D|\text{Graph }0)}\right)$$

Power PC
$$\Delta V_i = \alpha_{i\gamma(i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\prod_{\delta(V_k)=1}(1-V_k)[1-\prod_{\delta(V_j)=1}(1-V_j)])$$

Sequential Bayesian Theory
$$P(\vec{w}_{t+1}|D_t, m) = \int P(\vec{w}_{t+1}|\vec{w}_t)P(\vec{w}_t|D_t, m)d\vec{w}_t$$
$$P(\vec{w}_{t+1}|D_{t+1}, m) = \frac{P(D_{t+1}|\vec{w}_{t+1}, m) * P(\vec{w}_{t+1}|D_t, m)}{P(D_{t+1}|D_t)}$$

Conditional Probabilistic Contrast
$$\Delta P_i = P(E|i \cap Q) - P(E|\neg i \cap Q)$$

Novel Models

Bayesian Correlation Optimization
$$Pr(\rho|W) = \frac{Pr(W|\rho)Pr(\rho)}{Pr(W)} \text{ where } W = X_1 + X_2$$

Moving k-Window
$$V = \frac{1}{k}(N(C,E) + N(\neg C, \neg E)) - \frac{1}{k}(N(\neg C, E) + N(C, \neg E))$$

Win-Stay, Lose Shift
$$V_{i+1} = \begin{cases} V_i + \frac{1-V_i}{2} & V_i \geq 0, \delta(E) = \delta(C) \\ -.5 & V_i \geq 0, \delta(E) \neq \delta(C) \\ V_i - \frac{1+V_i}{2} & V_i < 0, \delta(E) \neq \delta(C) \\ .5 & V_i < 0, \delta(E) = \delta(C) \end{cases}$$

Rescorla-Wagner with decay
$$\Delta V_i = \mu\alpha_{i\gamma(C_i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\sum_{\delta(C_j)} V_j)$$
$$\text{where } \mu = e^{-\gamma n}$$

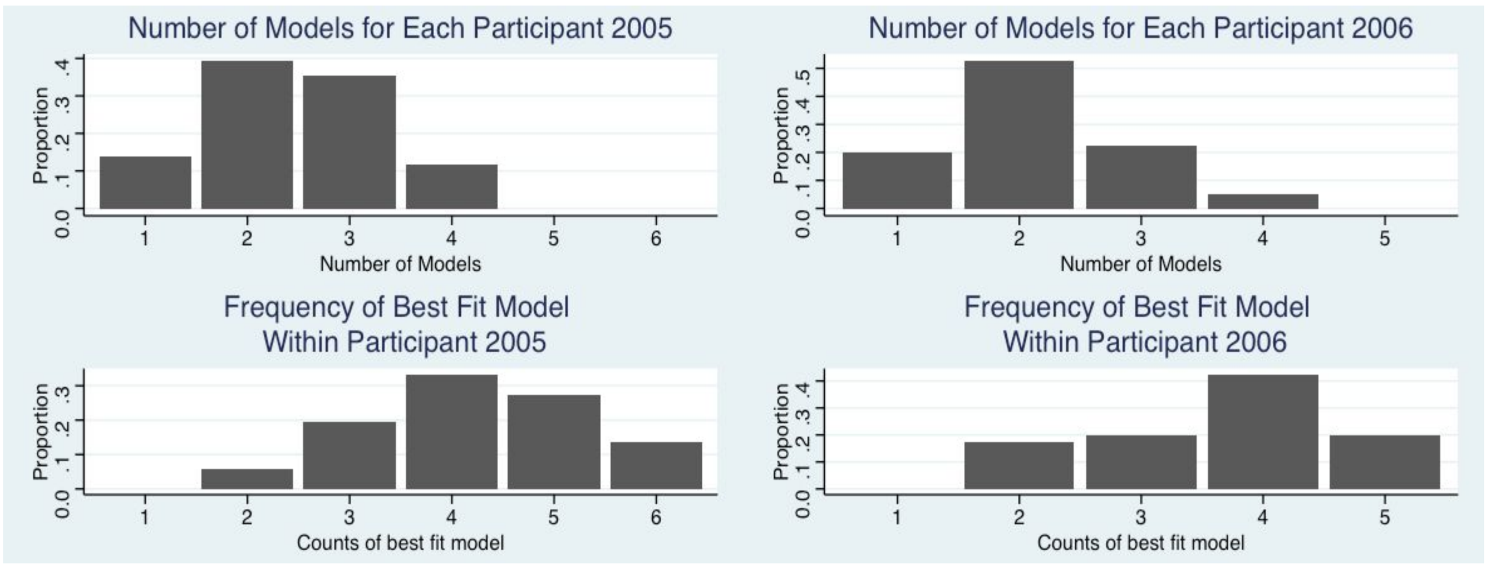
Rescorla-Wagner with stability of beliefs
$$\Delta V_i = \mu\alpha_{i\gamma(C_i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\sum_{\delta(C_j)} V_j)$$
$$\text{where } \mu = \max(\gamma\sigma, \frac{1}{\sqrt{n}})$$

Power PC with decay
$$\Delta V_i = \mu\alpha_{i\gamma(C_i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\prod_{\delta(C_k)=1}(1-V_k)[1-\prod_{\delta(C_j)}(1-V_j)])$$
$$\text{where } \mu = e^{-\gamma * n}$$

Power PC with stability of beliefs
$$\Delta V_i = \mu\alpha_{i\gamma(C_i)}\beta_{\delta(E)}(\lambda\delta(E) - \textstyle\prod_{\delta(C_k)=1}(1-V_k)[1-\prod_{\delta(C_j)}(1-V_j)])$$
$$\text{where } \mu = max(\gamma * \sigma, \frac{1}{\sqrt{n}})$$

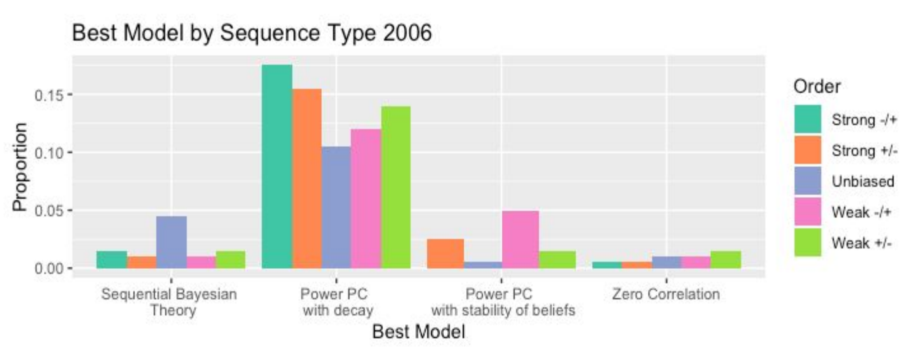
Results

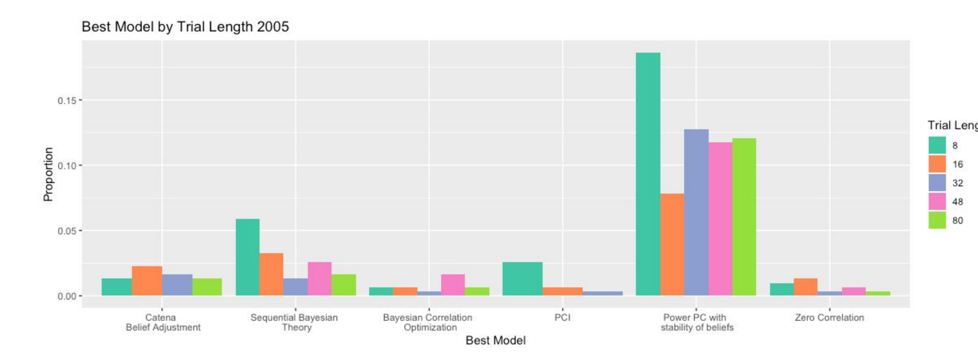
Below: Model consistency within subject

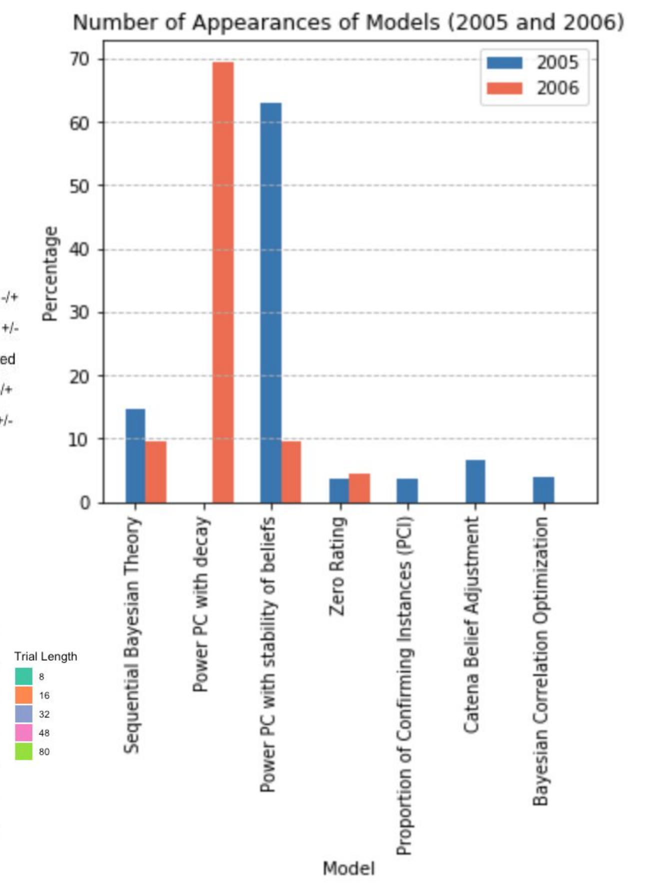


The histograms below show the models that appear more than 3% of the time

Below: Model breakdown by trial length and sequence type







Conclusions

- Overall, the expanded causal power theories exceeded all other causal learning models in prediction accuracy
- The Sequential Bayesian Theory performed better on unbiased sequences, plausibly due to the order invariance property of Bayesian models
- Both the Catena Belief Adjustment and PCI models generated improved predictions on shorter sequences, suggesting that these models capture the volatility of early causal predictions
- Future research directions include an analysis into model consistency within subject and the overlap in the prediction space of various models