CONJOINT ANALYSIS FOR CALIBRATING BAYES NET PARAMETERS

Stephen Walsh, Angie Dalton, Amanda White, Paul Whitney*

May 2010

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

What is a Bayes Net?

- Bayes nets (BN) are graphical representations of relationships among variables, provide generalized, quantitative modeling capability with established methods for integrating data, and compactly represent causal interactions in a complex environment where uncertainty predominates (Whitney, et al., 2009).

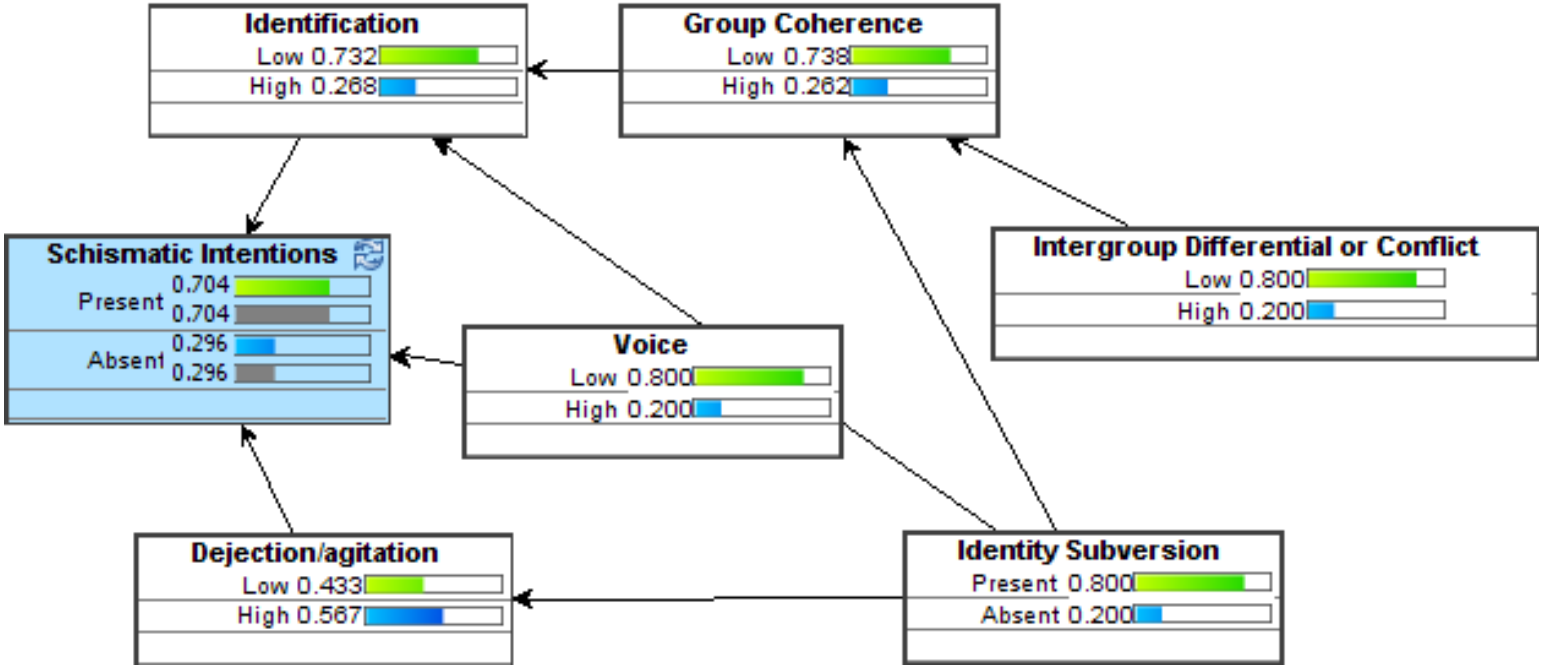
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}\,(X_i \mid \text{Parents}\,(X_i))$$

- In the most basic case, conditional distribution is represented as a conditional probability table (CPT) to give the distribution over X_i for each combination of parent values (AIIMA UCBerkeley, 2004)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Model Example: Group Schism



Bayesian network representation of mechanisms related to political group schism. The colored bars are representing the marginal probability distributions for each node given the current model parameters (source: Sani, 2005).

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Common Issues in Calibrating BN Parameters

- The "curse of dimensionality."
  - Small to moderately sized data sets may contain little to no observations in some regions even if the true probabilities are not extremely small.
  - Estimating the conditional probabilities as "0" based on a moderately sized data set can invoke a strong and possibly tenuous assumption.
- Complete lack of data on some model variables.
  - In practice it is common to encounter relevant extant data that cover some but not all of the model variables.
  - While possible to implement valid parameter estimation algorithms that help mitigate the issue, the unobserved variable's parameters cannot be calibrated (Walsh, et al., 2010).
- Expert elicitation is used but can be tedious (Woodbery, et al., 2005).

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Proposed Solution — Conjoint Analysis

- Conjoint analysis is commonly used in marketing research to elicit and model product attribute preferences from consumers.
- Rich theory and long history of empirical success (Luce and Tukey, 1964)
- Objective:
  - Employ conjoint analysis as a method for eliciting expert opinions and incorporating them into the model parameters.
- Offers ability to:

1. Handily aggregate the opinions of multiple experts into a single model calibration
2. Be considered a stand-alone approach to calibration if domain experts are the only available source of relevant information
3. Easily lend itself to calibrating subsets of the model if relevant data exist for the other subsets.
4. More readily employ the natural decision-making process of the human mind as opposed to explicitly asking the supposed value of the true conditional probability.

Conjoint Analysis Elicitation: User Interface

Expert Elicitation: Conjoint Analysis

? others providing input

For each pair of scenarios below, rate whether you think the left or right one is more likely.
This information will be used to update the BACH model.

Social Psychological Model of Schism in Groups

Model description:

This model is focused on group splitting. This component is directly inspired from F. Sani (2005): When Subgroups Secede: Extending and Refining the Social Psychological Model of Schism in Groups, Personality and Social Psychology Bulletin (PSPB), Vol 31(8):1074-1086.

Scenario 1		Scenario 2	
Schismatic Intentions	Present	Schismatic Intentions	Present
Intergroup Differential or Conflict	High	Intergroup Differential or Conflict	Low
Group Coherence	High	Group Coherence	Low

Scenario 1 more likely — Equally likely — Scenario 2 more likely

Submit Skip

Schismatic Intentions:

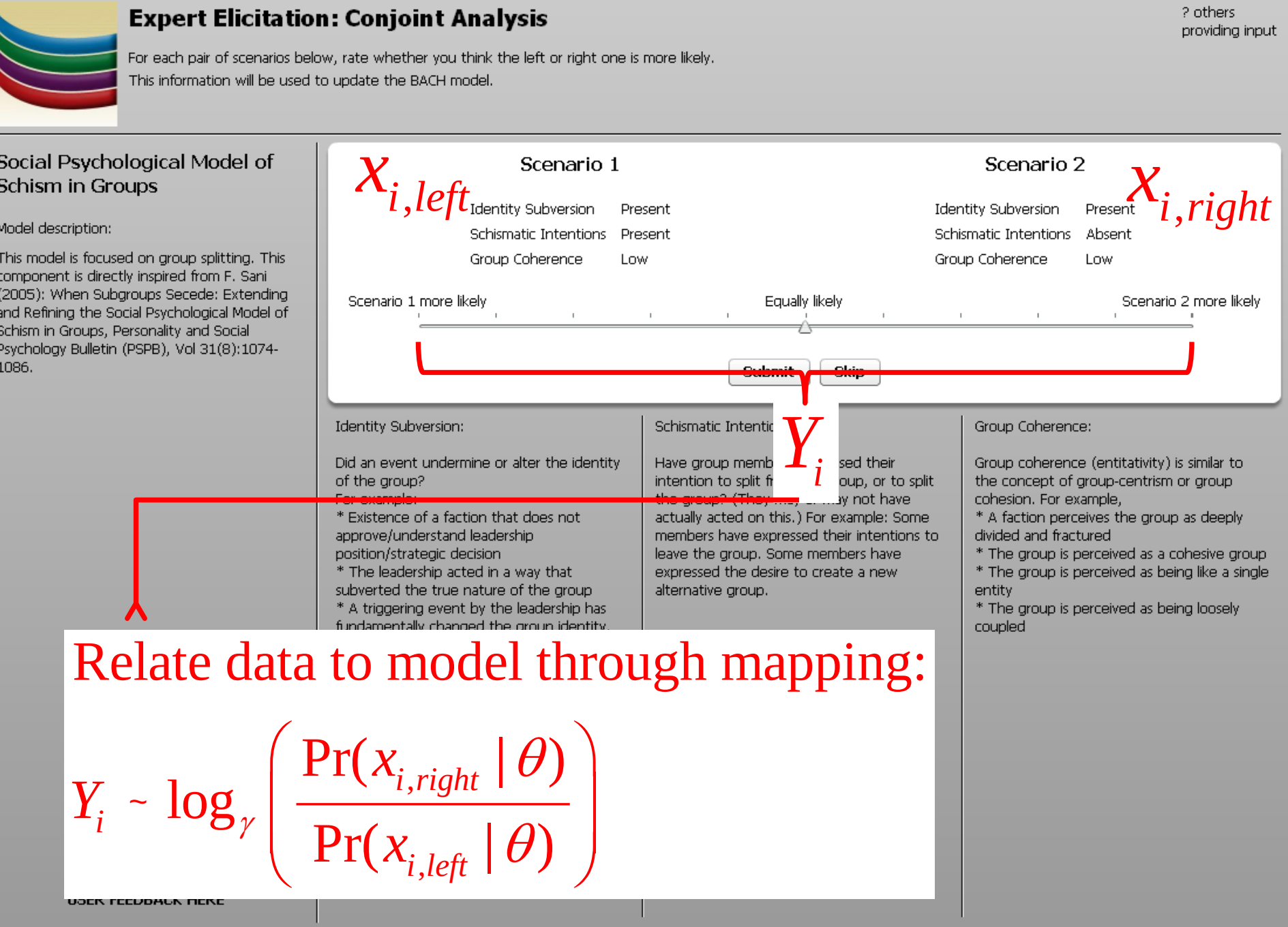
Have group members expressed their intention to split from the group, or to split the group? (They may or may not have actually acted on this.) For example: Some members have expressed their intentions to leave the group. Some members have expressed the desire to create a new alternative group.

Intergroup Differential or Conflict:

Is the group in competition with or in conflict with other similar groups?

Group Coherence:

Group coherence (entitativity) is similar to the concept of group-centrism or group cohesion. For example,
* A faction perceives the group as deeply divided and fractured
* The group is perceived as a cohesive group
* The group is perceived as being like a single entity
* The group is perceived as being loosely coupled



Relate data to model through mapping:

$$Y_i \sim \log_\gamma \left(\frac{\Pr(x_{i,right} \mid \theta)}{\Pr(x_{i,left} \mid \theta)} \right)$$

Modeling Expert Opinions

- Heuristic Objective:
  - Given expert opinions on conjoint scenarios, compute model parameters such that the model's perspective is "as close as possible" to the expert opinions.
- Formulate as optimization problem; compute

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \left[Y_i - \log_{\gamma} \left(\frac{\Pr(x_{i,right} \mid \theta)}{\Pr(x_{i,left} \mid \theta)} \right) \right]^2 + \lambda r(\theta)$$

where

Y_i : Expert Opinion

$x_{i,left}$, $x_{i,right}$: The left and right scenarios

$\Pr(X \mid \theta)$: The scenario probability

θ: The Bayes Net Parameters

$\log_{\gamma}$: A mapping from the likelihood ratio onto the range of Y

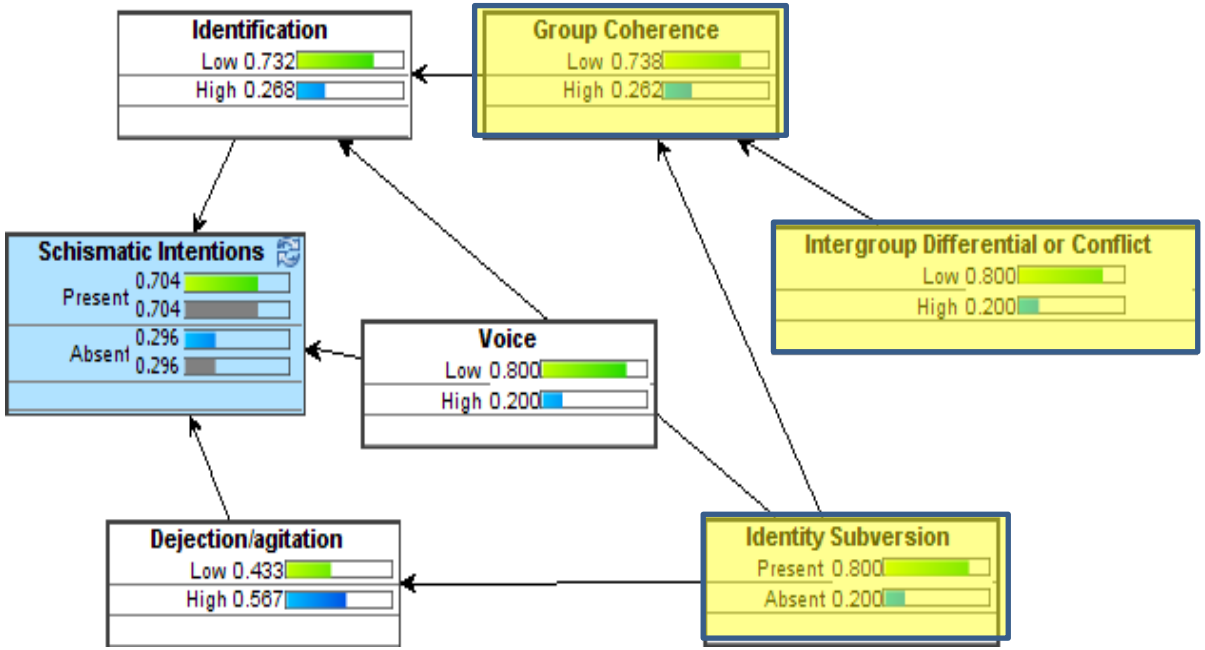
$r(\theta)$: Regularization function, assists computation when scenario coverage of model is poor

λ : Regularization parameter

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Verification

- Verify that calculations can amply incorporate opinions into model
- Simulate-recapture experiment:
  - Schism model is initialized with a parameter set
  - Simulate “expert opinion” data from known parameters, perturb with noise
  - Perform optimization, initialize optimization routine at naïve parameter set (uniform probabilities)
  - Does the calculation recapture the parameters whence the data was generated?



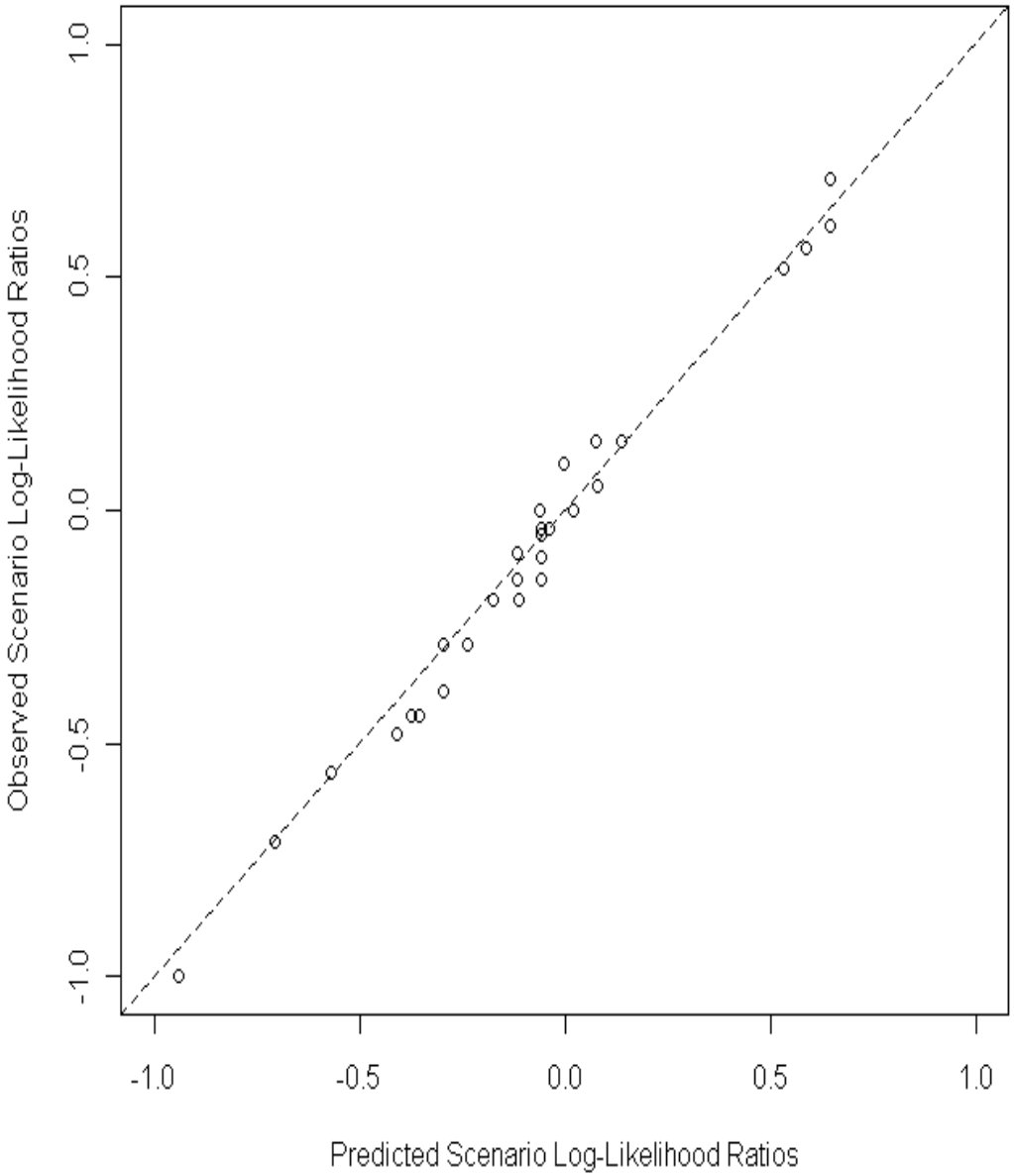
Considered conjoint scenarios from the yellow variables

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Verification results

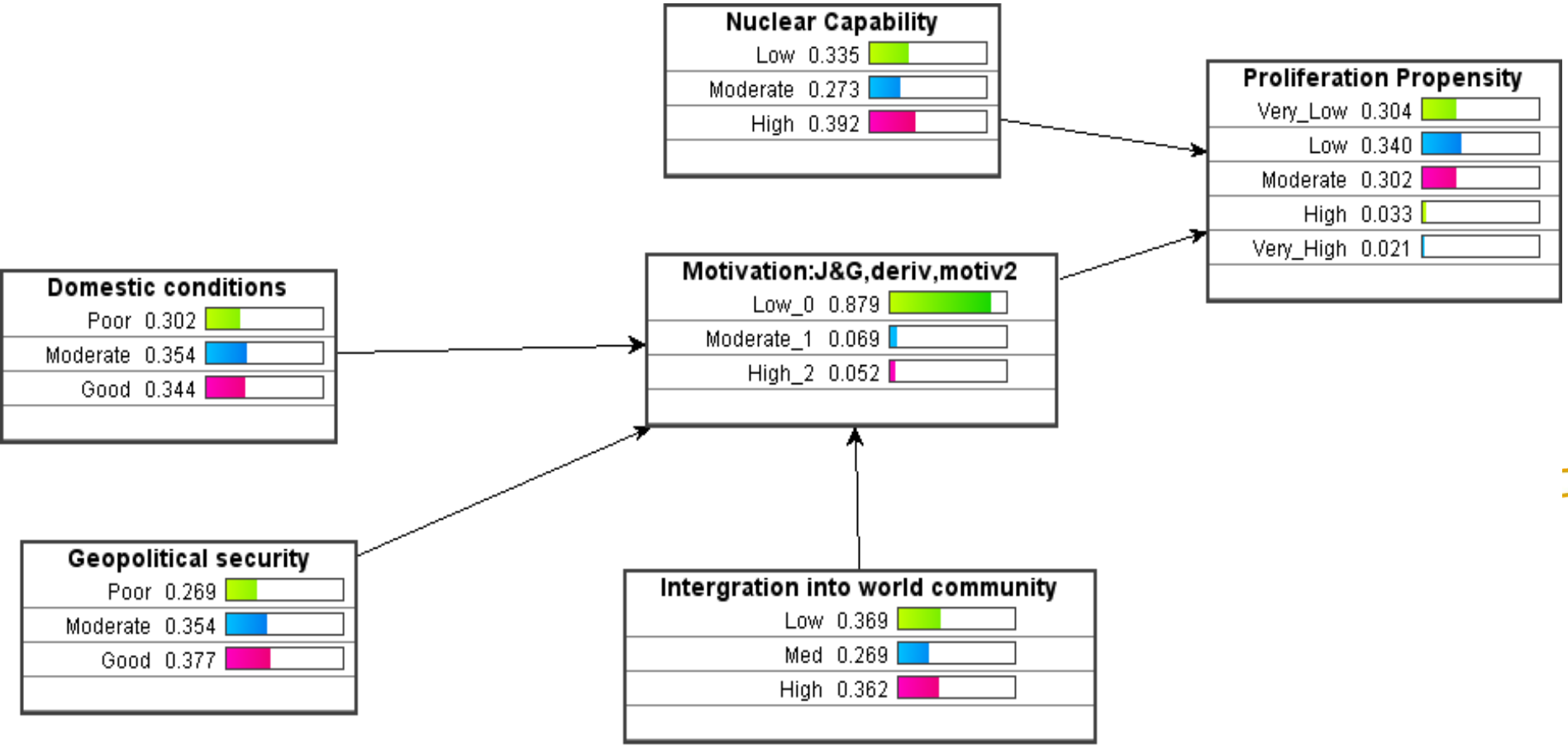
- Plot of Observed (simulated) Expert Opinion Data vs. Predicted values *after* model calibration

- Strong collinearity indicates successful incorporation of opinions into the model



Application: Nuclear Proliferation Model

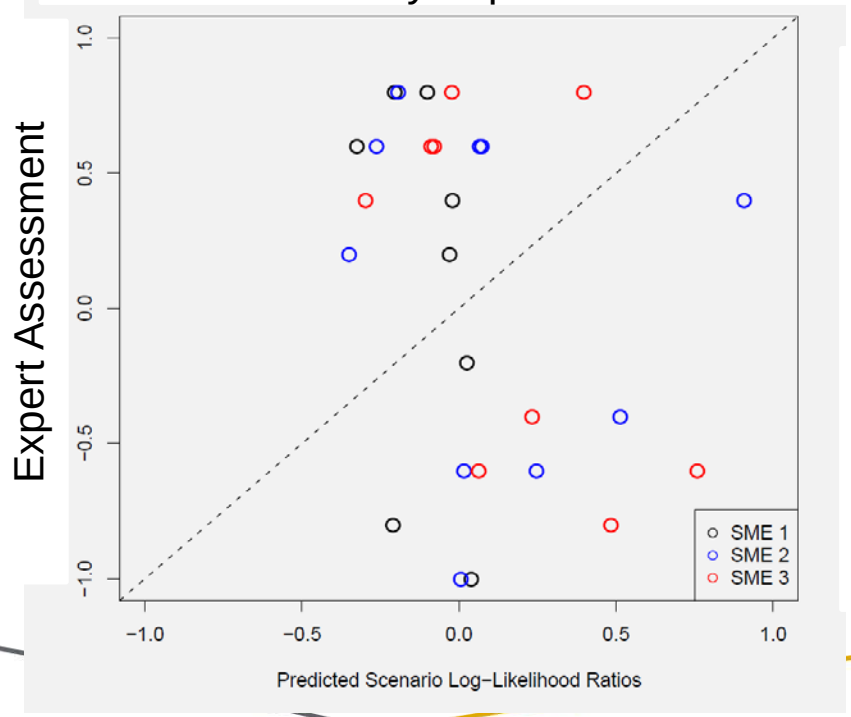
- Subset of larger model which describes the factors that contribute to nuclear proliferation
- Elicited 3 subject matter experts (SME) on their opinions of the relative likelihood of 30 conjoint scenarios



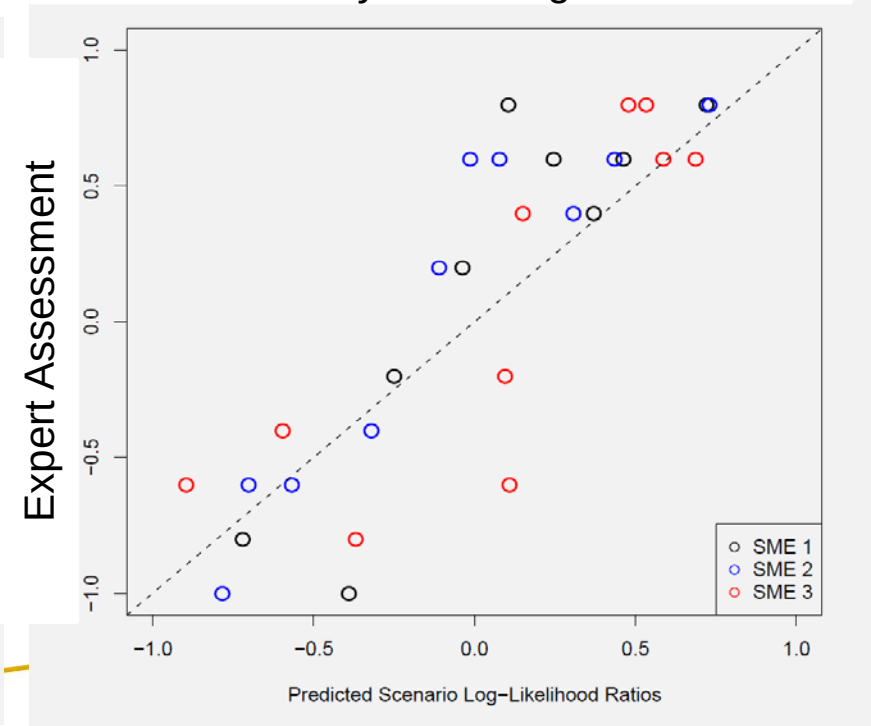
Application: Nuclear Proliferation Model

- Diagnostic plots indicate the incorporation of SME opinion into the model; the degree of calibration is optimized given the level of variability of the SME responses.

Original model – probability tables filled in by Expert



Calibrated model – probability tables filled in by PNNL algorithm



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Research Outcomes

- Successfully demonstrated the elicitation interface through which expert opinions were collected and measured.
- Presented initial work in developing a computational framework for translating the elicited opinions into the model parameters.
- Demonstrated that a model subset can be successfully calibrated from conjoint analysis data even while assuming no prior knowledge of the true distributions.

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Directions for Future Research

- Continue to develop and apply the computational framework
- Experiment with real experts and data
- Investigate the usefulness of different conjoint analysis presentation designs to improve BN model predictions
- Advance research in developing new methods in linking multiple expert elicitation techniques with computational modeling
- Expand the application of integrated modeling methodology to broader issue areas

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Project Publications

Whitney, Paul and Stephen Walsh, 2010. Calibrating Bayesian network Representations of Social-Behavioral Models. The 2010 International Conference on Social Computing, Behavior Modeling, and Prediction (SBP10). March 30, 2010. NIH Campus Bethesda, MD.

Walsh SJ, Dalton AC, Whitney PD, White AM. *Parameterizing Bayesian Network Representations of Social-Behavioral Models by Expert Elicitation: A novel application of conjoint analysis.* Workshop on Current Issues in Predictive Approaches to Intelligence and Security Analytics held in conjunction with IEEE Intelligence and Security Informatics Conference in Vancouver, BC, 23-26 May, 2010.

Key References

Gelman A, Carlin J, Stern H, Rubin D (2004) Bayesian Data Analysis, *2nd Edition, Chapman and Hall,* Boca Raton, FL.

Riggelsen, C. (2006) Learning Parameters of Bayesian Networks from Incomplete Data via Importance Sampling, International Journal of Approximate Reasoning , 42,69-83.

Jensen, F., Nielsen, T. (2007) Bayesian Networks and Decision Graphs, *2nd Edition, Springer, New York, NY.*

R: A Language and Environment for Statistical Computing.

gRain: A Graphical Independence Network Package in R.

R. Luce, and J. W. Tukey. "Simultaneous conjoint measurement: a new type of fundamental measurement," Journal of Mathematical Psychology, vol. 1, pp 1-27, 1964.