George Mason University (Mason)
  • Charles Twardy (PI), Kathryn Laskey (Co-PI)
  • Robin Hanson, Wei Sun, Max Tsvetovat, Don Gantz
Australian Center for Excellence in Risk Analysis
  • Mark Burgman
  • Fiona Fidler, Neil Thomason, Ann Nicholson, Aidan Lyon (U.MD)
James Madison University
  • Noel Hendrickson
Mercyhurst College
  • Kristan Wheaton
nemoSibi Ltd. (Software Platform)
  • Dave Perry
KaDSCi LLC (Decision Analysis)
  • Dan Maxwell

And a supporting cast
The problem: how best to aggregate?

- For prediction, aggregates should outperform individuals.
  - *They do.*
- Weighted aggregates should outperform unweighted.
  - *They don’t.*
- Why?
  - Flat Maximum? (von Winterfeldt and Edwards)
    - But then why so much room between experts and statistical models?
  - Community hasn’t found strong factors to weight
    - Training, experience, confidence, and prestige: *no*
    - Tetlock’s thinking style: *yes, but*
  - Past performance might
    - Analysts resist measuring performance (Kent, Heuer, Johnston)
- ACE: How to improve on the unweighted average?
Schematic View of Our Approach

Question

Orderly Transition
Kim Jong Il Remains in Power
Revolt
Natural Causes

Date Leader Leaves
2011
J F M A M J J A S O N D
Our Key Ideas:

- **Problem decomposition:**
  - Contingent to the core
  - Estimate better

- **Advanced Elicitation:**
  - Better individual estimates
  - Counter biases
    - Group-think
    - Anchoring
    - Halo
    - Overconfidence

- **Bayesian Combo Exchanges:**
  - Prediction Exchanges
  - Conditional Forecasts
  - Bayesian Updating

- **Learning/Analysis:**
  - Pools: help weight forecasts
  - Markets: autotraders

- **A Diverse Analyst Pool**
  - Diversity trumps ability
Decomposition

(Related: Conditional, Combinatorial)

<table>
<thead>
<tr>
<th>Kim Jong Il Remains in Power</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>85.0</td>
</tr>
<tr>
<td>No</td>
<td>15.0</td>
</tr>
</tbody>
</table>

The question to be answered is "What is the probability that Kim Jong Il remains continuously in power as the Supreme Leader of North Korea through 11:59 P.M. GMT the 31st of December 2011?"
Decomposed Model

What is the probability that Kim Jong Il will die of natural causes before 11:59 GMT December 31st 2011?

What is the probability that Kim Jong Il will willingly transition power to his son (or some other person) before 11:59 GMT December 31st 2011?

What is the probability that a revolt will unseat Kim Jong Il from power before 11:59 GMT December 31st 2011?

Is there another reason Kim Jong Il could leave power as the Supreme Leader before 11:59 GMT December 31st 2011?
Prediction Exchange Team

- **17 years of domain expertise**
  - Helped spark the collective intelligence industry in 1994 by developing the original prediction market software
  - Responsible for many of the innovations that are commonplace today

- **Extensible software platform**
  - Easy to use and administer
  - Highly customizable and configurable
  - Comprehensive API facilitates integration with other systems
  - Robust and secure hosting environment

- **Deployment experience**
  - Popularized the commercial use of collaborative forecasting and led many large-scale projects for public and private organizations, such as: General Electric, Motorola, Bank of America, Lockheed Martin, Best Buy, General Mills, UnitedHealth, and the Missile Defense Agency.
## 2008 US President Example

*From InTrade.com*

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Nominate?</th>
<th>Win?</th>
<th>Win if Nom.?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>74.3-76.0%</td>
<td>46.4-47.4%</td>
<td>61-64%</td>
</tr>
<tr>
<td>Clinton</td>
<td>12.1-12.4%</td>
<td>6.6-7.7%</td>
<td>53-64%</td>
</tr>
<tr>
<td>Gore</td>
<td>1.5-1.8%</td>
<td>1.6-1.7%</td>
<td>89-100%</td>
</tr>
<tr>
<td>McCain</td>
<td>96.1-96.2%</td>
<td>37.8-38.4%</td>
<td>39-40%</td>
</tr>
<tr>
<td>Giuliani</td>
<td>1.3-1.4%</td>
<td>0.2-0.4%</td>
<td>14-31%</td>
</tr>
<tr>
<td>Paul</td>
<td>1.0-1.1%</td>
<td>0.6-0.7%</td>
<td>54-70%</td>
</tr>
</tbody>
</table>
Imagine A Dashboard

![Image](image_url)

Ave. Score: **12,459**

<table>
<thead>
<tr>
<th></th>
<th>Us</th>
<th>Them A</th>
<th>Them B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$320</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>May ’09</td>
<td>Mar ’09</td>
<td>July ’09</td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozoop 38%</td>
<td>69%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo 59%</td>
<td>8%</td>
<td>43%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total 120K</td>
<td>148K</td>
<td>97K</td>
</tr>
<tr>
<td></td>
<td>Base model 82K</td>
<td>65K</td>
<td>88K</td>
</tr>
<tr>
<td></td>
<td>Via internet 43K</td>
<td>12K</td>
<td>73K</td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine $30K</td>
<td>$50K</td>
<td>$3K</td>
</tr>
<tr>
<td></td>
<td>Circulars $45</td>
<td>$34K</td>
<td>$39K</td>
</tr>
<tr>
<td></td>
<td>Us</td>
<td>Them A</td>
<td>Them B</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$320</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>May ’09</td>
<td>Mar ’09</td>
<td>July ’09</td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozeep 38%</td>
<td>69%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo 59%</td>
<td>8%</td>
<td>43%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total</td>
<td>120K</td>
<td>148K</td>
</tr>
<tr>
<td></td>
<td>Base model</td>
<td>82K</td>
<td>65K</td>
</tr>
<tr>
<td></td>
<td>Via internet</td>
<td>43K</td>
<td>12K</td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine</td>
<td>$30K</td>
<td>$50K</td>
</tr>
<tr>
<td></td>
<td>Circulars</td>
<td>$45</td>
<td>$34K</td>
</tr>
<tr>
<td></td>
<td>Us</td>
<td>Them A</td>
<td>Them B</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$320</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>May’09</td>
<td>Mar ’09</td>
<td>July ’09</td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozoop</td>
<td>42%</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo</td>
<td>59%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total</td>
<td>120K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Base model</td>
<td>82K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Via internet</td>
<td>43K</td>
<td>12K</td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine</td>
<td>$30K</td>
<td>$50K</td>
</tr>
<tr>
<td></td>
<td>Circulars</td>
<td>$45</td>
<td>$34K</td>
</tr>
</tbody>
</table>

If We Have Autozoop, you gain 53. But if We Don’t Have It You lose 78. OK?
Make an Assumption

<table>
<thead>
<tr>
<th></th>
<th>Us</th>
<th>Them A</th>
<th>Them B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$253</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>Apr ’09</td>
<td>Mar ’09</td>
<td><strong>Assume Mar</strong></td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozoop 38%</td>
<td>69%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo 59%</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total 120K</td>
<td>148K</td>
<td>107K</td>
</tr>
<tr>
<td></td>
<td>Base model 82K</td>
<td>65K</td>
<td>94K</td>
</tr>
<tr>
<td></td>
<td>Via internet 43K</td>
<td>12K</td>
<td>84K</td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine $30K</td>
<td>$50K</td>
<td>$17K</td>
</tr>
<tr>
<td></td>
<td>Circulars $45</td>
<td>$34K</td>
<td>$49K</td>
</tr>
</tbody>
</table>

**Scenario: 15%**

**Ave. Score: 10,724**
Add 2nd Assumption

<table>
<thead>
<tr>
<th></th>
<th>Us</th>
<th>Them A</th>
<th>Them B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$253</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>Apr ’09</td>
<td>Mar ’09</td>
<td>Assume Mar</td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozoop</td>
<td>38%</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo</td>
<td>59%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total</td>
<td>185K</td>
<td>148K</td>
</tr>
<tr>
<td></td>
<td>Base model</td>
<td>97K</td>
<td>65K</td>
</tr>
<tr>
<td></td>
<td>Via internet</td>
<td>78K</td>
<td>12K</td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine</td>
<td>Assume $40K</td>
<td>$50K</td>
</tr>
<tr>
<td></td>
<td>Circulars</td>
<td>$45</td>
<td>$34K</td>
</tr>
<tr>
<td></td>
<td>Us</td>
<td>Them A</td>
<td>Them B</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Base Price</strong></td>
<td>$240</td>
<td>$187</td>
<td>$253</td>
</tr>
<tr>
<td><strong>Ship Date</strong></td>
<td>Apr ’09</td>
<td>Mar ’09</td>
<td><strong>Assume Mar</strong></td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Autozoop</td>
<td>42%</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Fizzywoo</td>
<td>59%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Unit Sales</strong></td>
<td>Total</td>
<td>185K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Base model</td>
<td>97K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Via internet</td>
<td>78K</td>
<td></td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td>Magazine</td>
<td>Assume $40K</td>
<td>$50K</td>
</tr>
<tr>
<td></td>
<td>Circulars</td>
<td>$45</td>
<td>$34K</td>
</tr>
</tbody>
</table>

- If we have Autozoop, you gain 40
- But if we don’t have it, you lose 62. OK?
Editing Interface Is Transparent

I directly change the consensus

If my edit increases the consensus chance of true state, I win. If decreases, I lose.

Consensus

My Edits

My Score

Truth
Elicitations: Expert Judgment & Risk

- Expert judgement
- Spatial analysis
- Stakeholder mapping
- Consequences
- Biosecurity intelligence
- Disease freedom/eradication
- Where should we monitor/search?

Social networks Intelligence software

Maxent

Bayes nets

Inspection / searching: cost-effectiveness analysis

16
MCIIS Overview

• 19 Years
• 350 Student-Analysts
• 12 Full-time Faculty, *Countless Adjuncts*
• Graduates work in Business, Law Enforcement, and National Security
  • …And Internationally

• High placement in the IC
• Network
• Professional degree
Study Participants:

- 60+ Students in the Information Analysis Program (a Undergraduate Major for Future Intelligence Analysts that Focuses on Analytic Methodology) Many of Whom Also Have Specialty Subject Area Knowledge (e.g. East Asia, Middle East, etc.)

- Diverse SME Faculty Pool

Research in Elicitation Methodology:

- Institute for National Security Analysis (Research Institute That Works to Discover, Develop, and Deliver Analytic Methods for Intelligence and National Security) With a Special Focus on Cognitive Methods/Critical Thinking/Reasoning.

- Planned Research: How Counterfactual Reasoning and Systems Dynamics Can Be Adapted to Help Elicit Best Responses Through Improved Mental Structuring of the Question and Its Potential Answers
Data-Driven Performance Analysis

- IARPA INPUT
- Qs
- AUX INPUT
- DEMOGRAPHICS
- DATA
- INTERFACE MODULE
- FEEDBACK
- DATA DRIVEN MODELING MODULE
- DECISION
## Milestones

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions for Problem Set 1</td>
<td>1</td>
</tr>
<tr>
<td>IRB Approval, 100 Participants, Software V1</td>
<td>3</td>
</tr>
<tr>
<td>Site Visit &amp; Web Portal</td>
<td>4</td>
</tr>
<tr>
<td>Static BN PM elicitation</td>
<td>6</td>
</tr>
<tr>
<td>500 Participants</td>
<td>7</td>
</tr>
<tr>
<td>Site Visit 2: Comb’l UI; 15% &gt; ULinOP</td>
<td>9</td>
</tr>
<tr>
<td>At least one manuscript for publication</td>
<td>10</td>
</tr>
<tr>
<td>Year-end report, Milestone Y2</td>
<td>11</td>
</tr>
</tbody>
</table>
Y1 Timeline

<table>
<thead>
<tr>
<th>Task</th>
<th>2010</th>
<th>Qtr 1 2011</th>
<th>Qtr 2 2011</th>
<th>Qtr 3 2011</th>
<th>Qtr 4 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) IARPA ACE Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Elicitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1) Structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2) PM-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3) Interval/Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1) Standard Prediction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2) Develop PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3) PM elicitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4) Static PM elicitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5) Raw Combinatorial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5.1) PM elicitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6) Text mining side</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7) Year 1 Weighted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1) Demonstrate baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2) Establish software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3) Deliver V1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4) Customized front-end</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5) Provide software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.6) Customized development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.7) Preview combinatorial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.8) Export data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.9) Demonstrate software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.10) Ongoing support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Analysis &amp; Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Analyst Pool</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dependencies & Risks

- HSRB approval
- IARPA questions must pass the clarity test
- Correlations & information leakage among the performers and MITRE
- Experimental design and pool quality
Clarity of Questions

• IARPA Questions pass the clarity test.
  • We have not found it straightforward
  • Suggestion: Each team should have to check off on each question?

• Still, some Questions will be overtaken by events we didn't consider. We need a decision procedure.

• Suggestion:
  • If an untoward event happens, MITRE + group votes.
  • Does it now fail the clairvoyance test?
  • N-1 groups + MITRE to agree
  • MITRE suggests an outside panel
  • …
Correlation & Information Leakage

• Significant risk to evaluation
• Even a handful can correlate
  • Arbitrage
  • “Why are you using that other format?”
• Good for effectiveness, but bad for bake-off
• Alternatives:
  • Robust identity checking + hope
  • Teams obscure results: suboptimal
  • Publish all day-old estimates + change evaluation
Separate Pools

• Experiment has at least two variables:
  • Quality of expert pools
  • Quality of techniques

• If the pools *don’t* correlate, quality of the pool may dominate – a serious confound
  • A possible argument for info sharing

• MITRE’s new T&E: reserve 20 Qs for team ULinOP
  • Are there even more effective ways?
  • Controlled follow-up tests with random assignment?
  • Better ideas?
Conclusion

• Bayesian Combinatorial Markets
  • Software: Mason, nemoSibi  
  • Bayesian: Mason, KaDSCi 
  • Decomposition: KaDSCi, JMU, Mercyhurst
  • Elicitation: ACERA
  • Participants: JMU, Mercyhurst
    – TRIG, Mason, Open Recruitment
  • Data Analysis: Mason, KaDSCi