IARPA ACE: Mason Team
Decomposition-Based Aggregation
Daggre Forecasting Project

- **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

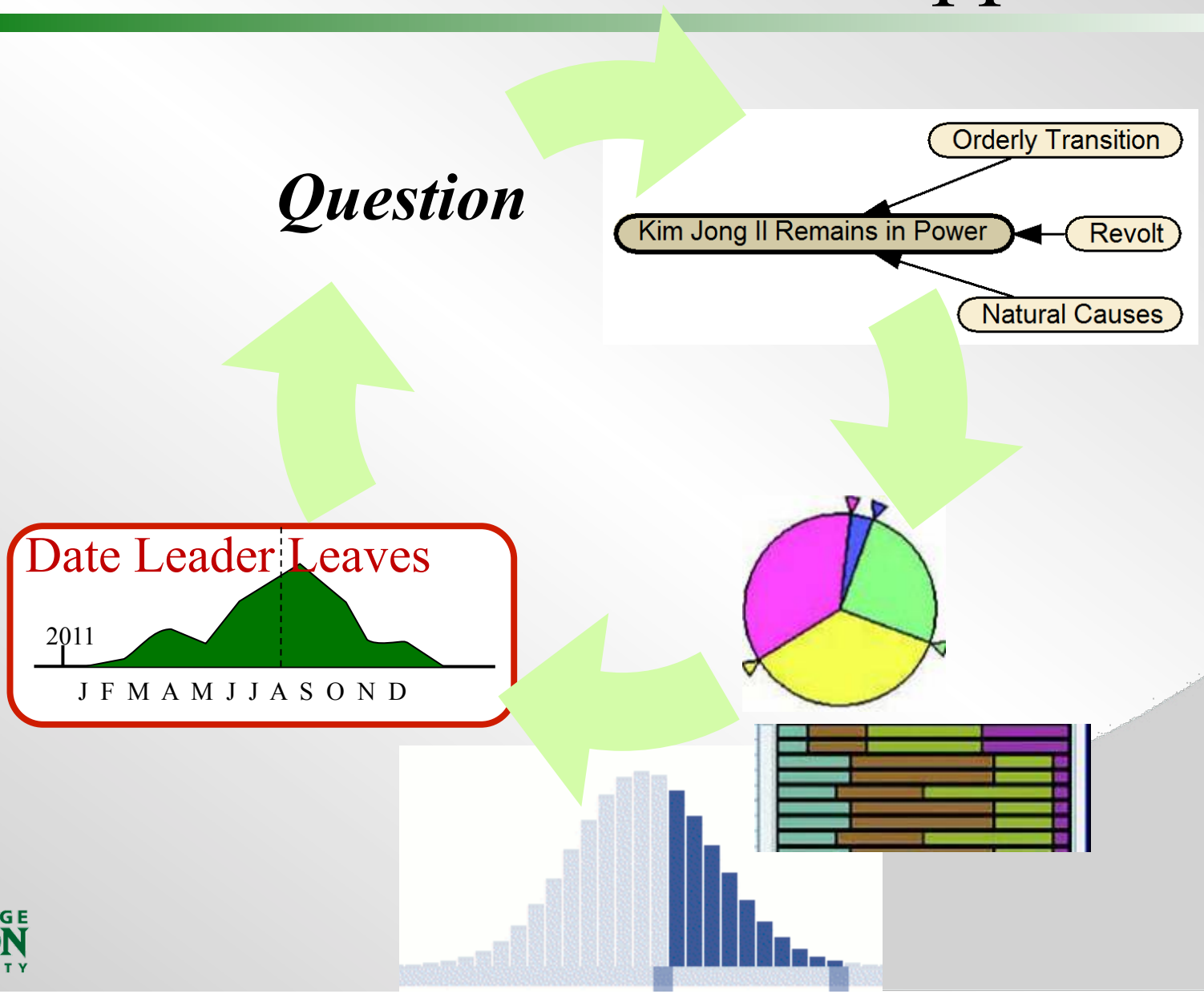
And a supporting cast

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?

GEORGE MASON UNIVERSITY

Schematic View of Our Approach



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

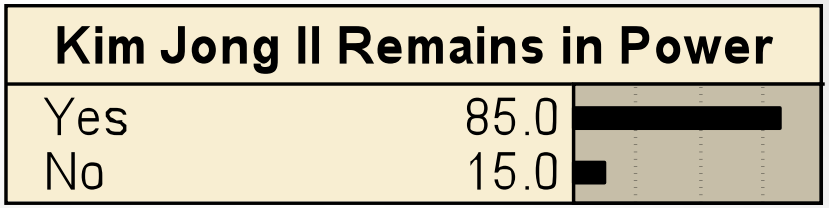
- **Learning/Analysis:**
- Pools: help weight forecasts
- Markets: autotraders

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

GEORGE MASON UNIVERSITY

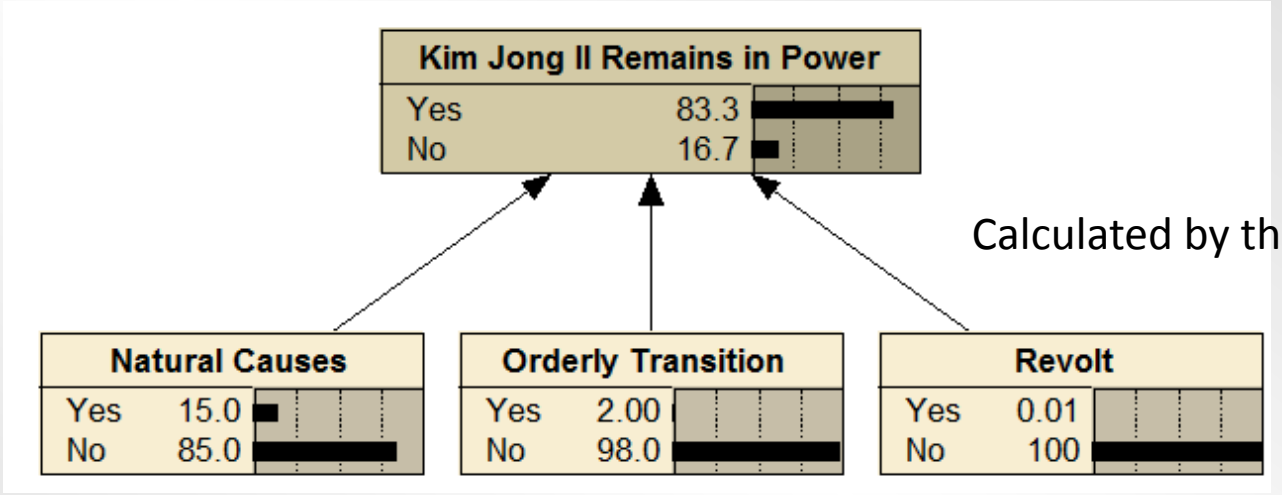
Decomposition

(Related: Conditional, Combinatorial)



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



Calculated by the model

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

2 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?

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

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

Prediction Exchange Team

nemoSibi

- **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.

GEORGE MASON UNIVERSITY

2008 US President Example

nemoSibi

From InTrade.com

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

GEORGE MASON UNIVERSITY

Imagine A Dashboard

nemoSibi

Ave. Score: 12,459

		Us	Them A	Them B
Base Price		$240	$187	$320
Ship Date		May ’09	Mar ’09	July ’09
Features	Autozoop	38%	69%	15%
	Fizzywoo	59%	8%	43%
Unit Sales	Total	120K	148K	97K
	Base model	82K	65K	88K
	Via internet	43K	12K	73K
Promotion	Magazine	$30K	$50K	$3K
	Circulars	$45	$34K	$39K

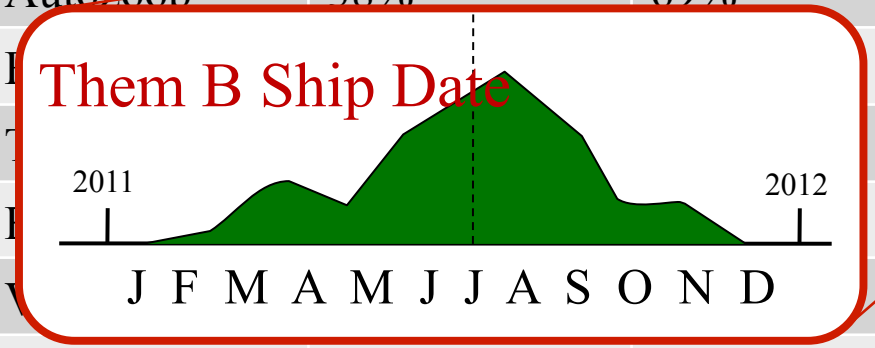
GEORGE MASON UNIVERSITY

Ask For Detail

nemoSibi

Ave.Score: 12,459

		Us	Them A	Them B
Base Price		$240	$187	$320
Ship Date		May ’09	Mar ’09	July ’09
Features	Autozoom	38%	69%	15%
				43%
Unit Sales				97K
				88K
				73K
Promotion	Magazine	$30K	$50K	$3K
	Circulars	$45	$34K	$39K



GEORGE MASON UNIVERSITY

Make An Edit

nemoSibi

Ave. Score: 12,459

		Us	Them A	Them B
Base Price		$240	$187	$320
Ship Date		May ’09	Mar ’09	July ’09
Features	Autozoop	42%	69%	15%
	Fizzywoo	59%	8%	42%
Unit Sales	Total	120K		
	Base model	82K		
	Via internet	43K	12K	73K
Promotion	Magazine	$30K	$50K	$3K
	Circulars	$45	$34K	$39K

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

GEORGE MASON UNIVERSITY

Make an Assumption

nemoSibi

Scenario: 15%

Ave. Score: 10,724

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

GEORGE MASON UNIVERSITY

Add 2nd Assumption

nemoSibi

Scenario: 2.3%

Ave. Score: 10,982

		Us	Them A	Them B
Base Price		$240	$187	$253
Ship Date		Apr ’09	Mar ’09	*Assume Mar*
Features	Autozoop	38%	69%	4%
	Fizzywoo	59%	8%	13%
Unit Sales	Total	185K	148K	107K
	Base model	97K	65K	94K
	Via internet	78K	12K	84K
Promotion	Magazine	*Assume $40K*	$50K	$17K
	Circulars	$45	$34K	$49K

GEORGE MASON UNIVERSITY

Edit As Before

nemoSibi

Scenario: 2.3%

Ave. Score: 10,724

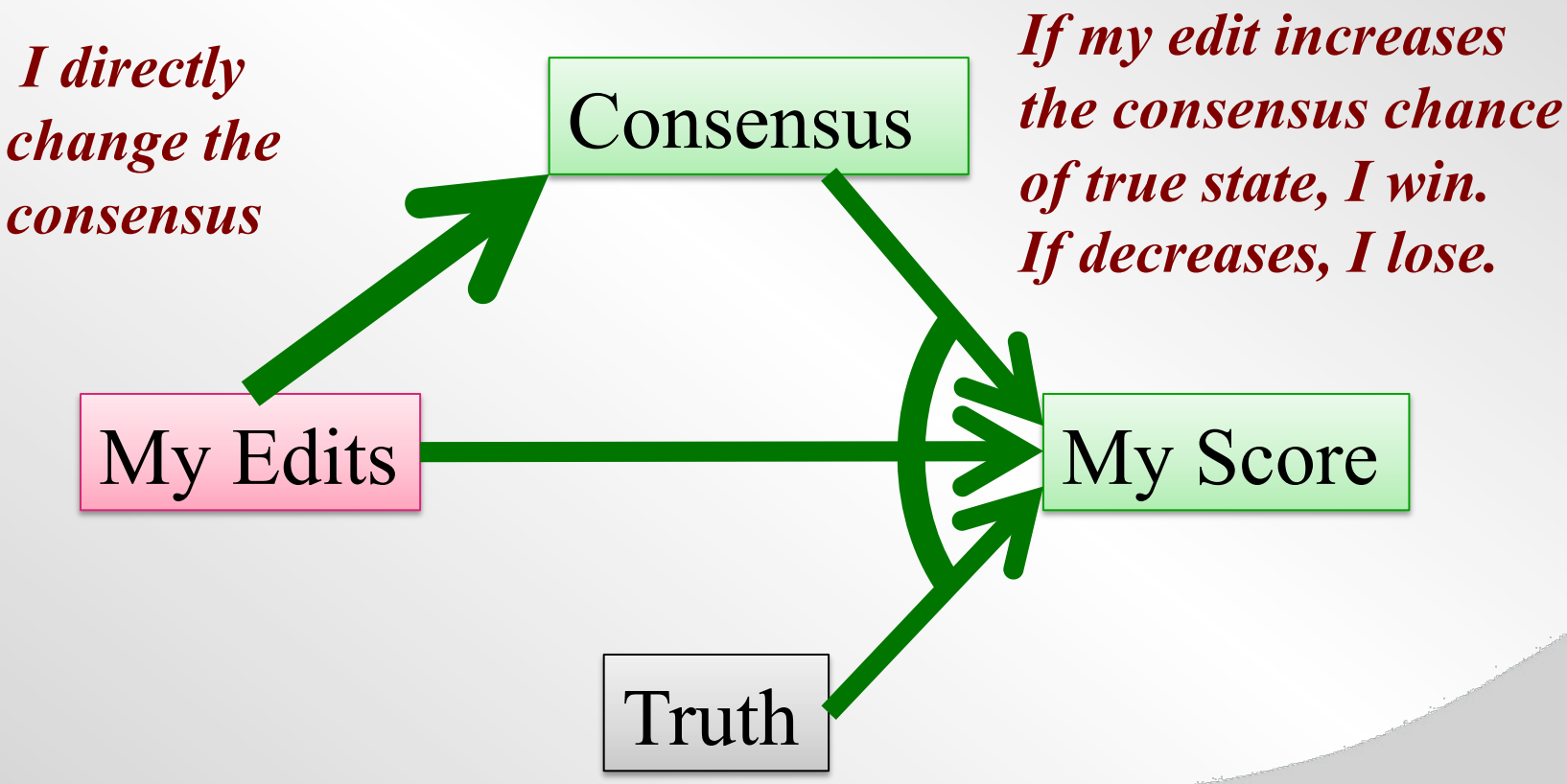
		Us	Them A	Them B
Base Price		$240	$187	$253
Ship Date		Apr ’09	Mar ’09	*Assume Mar*
Features	Autozoop	42%	69%	4%
	Fizzywoo	59%	8%	13%
Unit Sales	Total	185K		
	Base model	97K		
	Via internet	78K	12K	34K
Promotion	Magazine	*Assume $40K*	$50K	$17K
	Circulars	$45	$34K	$49K

If we have Autozoop,
you gain 40
But if we don’t have it
You lose 62. OK?

GEORGE MASON UNIVERSITY

Editing Interface Is Transparent

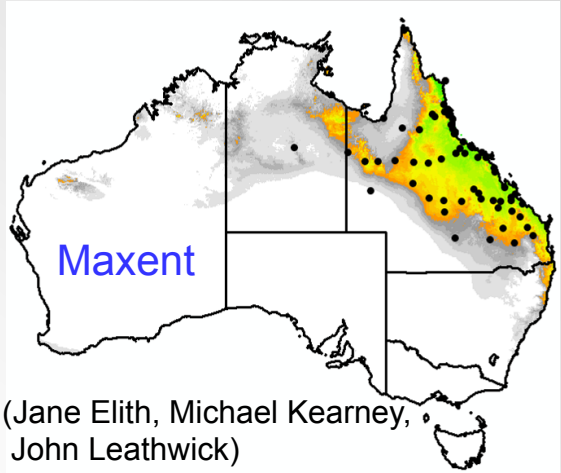
nemoSibi



Elicitations: Expert Judgment & Risk

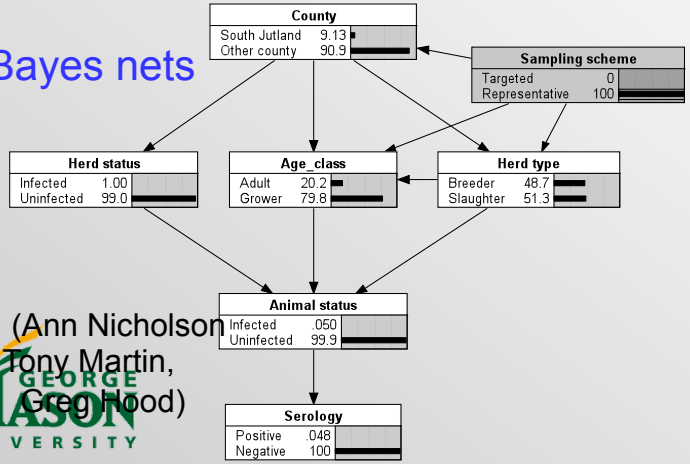
- Expert judgement
- Spatial analysis
- Stakeholder mapping
- Consequences
- Biosecurity intelligence
- Disease freedom/eradication
- Where should we monitor/ search?

Social networks
Intelligence software

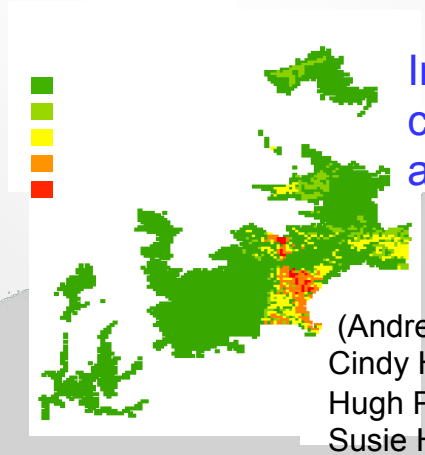


(Jane Elith, Michael Kearney, John Leathwick)

Bayes nets



(Ann Nicholson, Tony Martin, Greg Hood)



Inspection / searching: cost-effectiveness analysis

(Andrew Robinson, Rob Cannon, Cindy Hauser, Mick McCarthy, Hugh Possingham, Tracy Rout, Susie Hester, Oscar Cacho)

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

JMU INSA:

Institute for National Security Analysis ***Information Analysis Program***

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.)

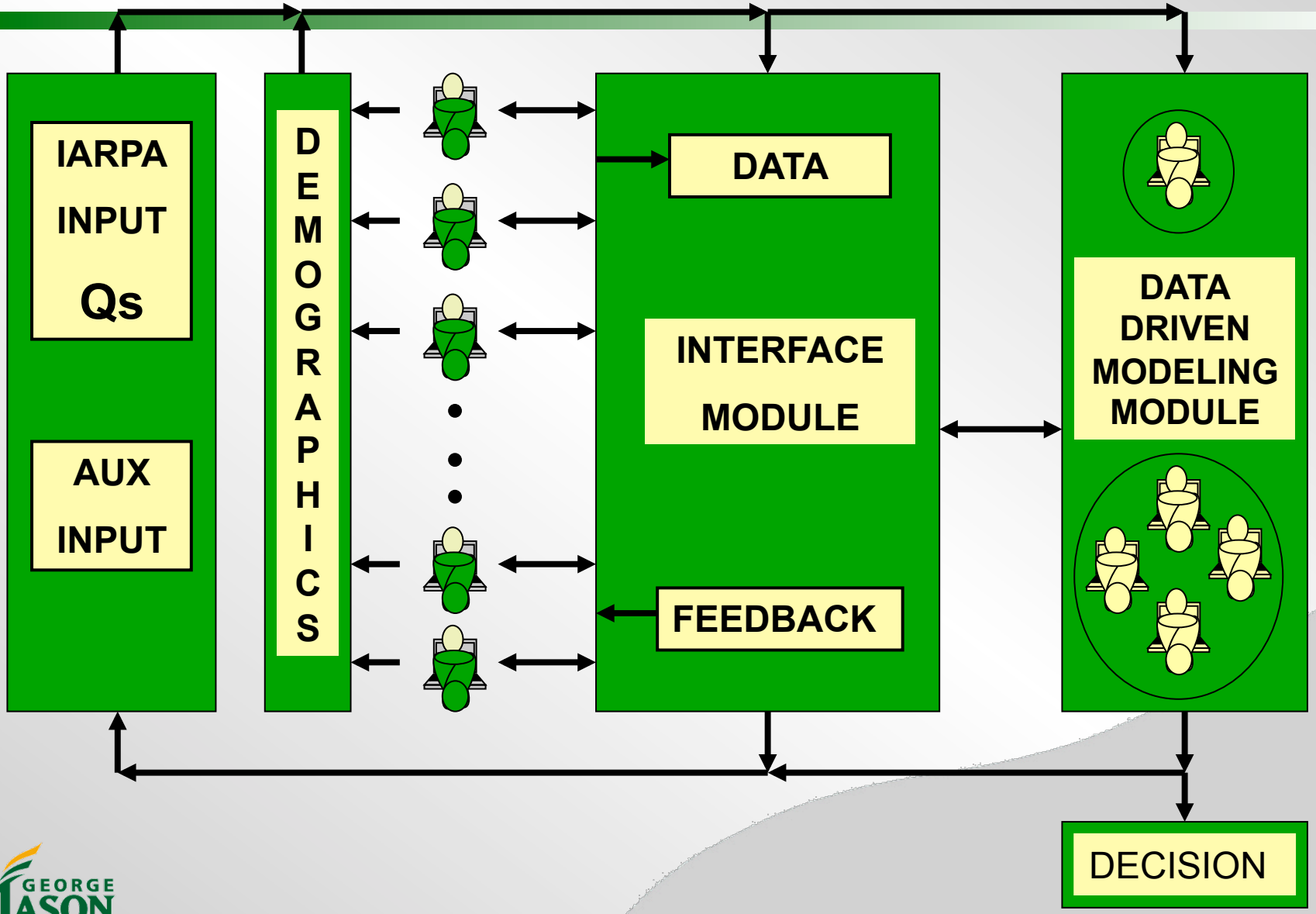
- Diverse SME Faculty Pool

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

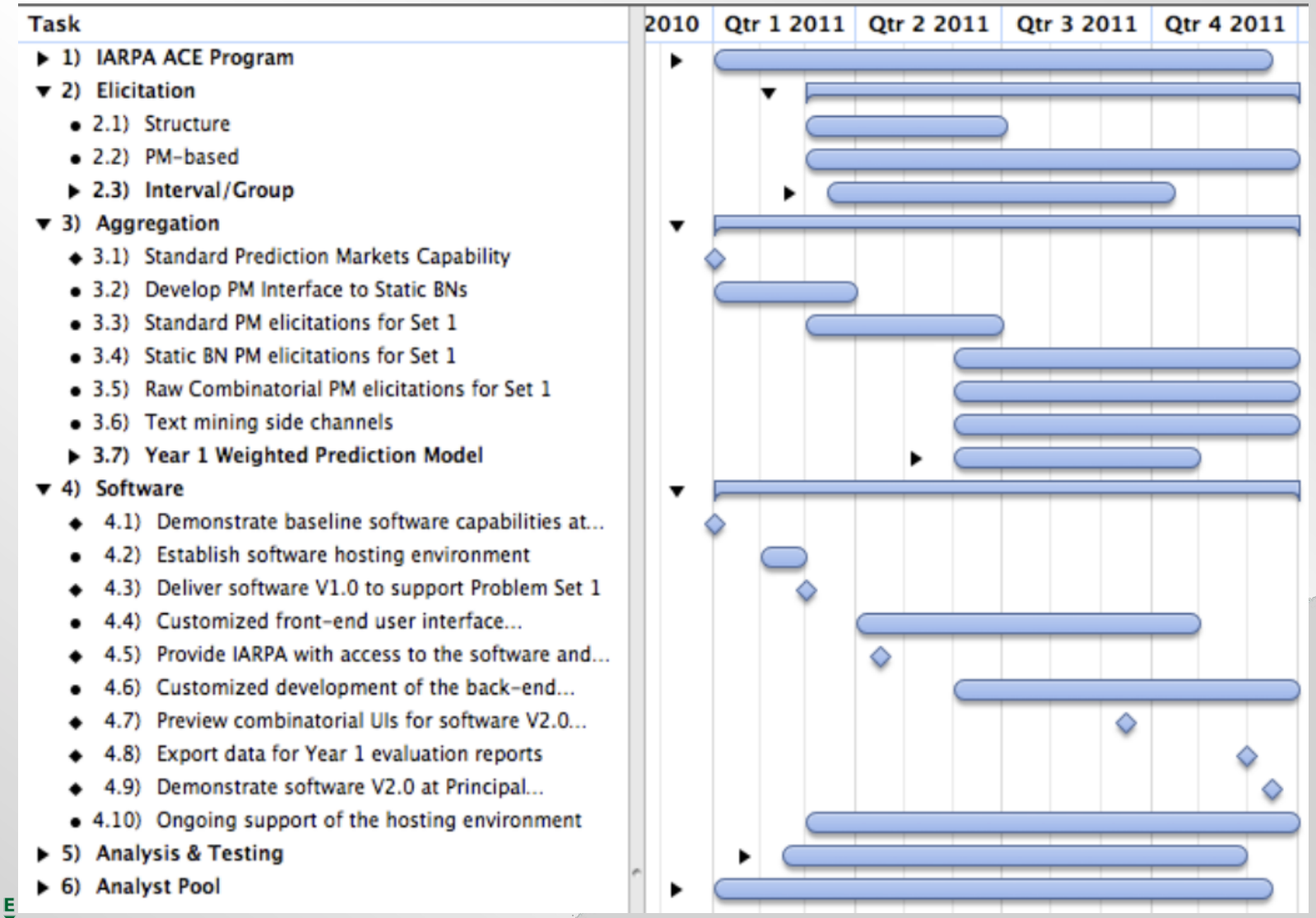


Milestones

Milestone	Month
Questions for Problem Set 1	1
IRB Approval, 100 Participants, Software V1	3
Site Visit & Web Portal	4
Static BN PM elicitations	6
500 Participants	7
Site Visit 2: Comb'l UI; 15% > ULinOP	9
At least one manuscript for publication	10
Year-end report, Milestone Y2	11

GEORGE MASON UNIVERSITY

Y1 Timeline



GEORGE MASON UNIVERSITY

Dependencies & Risks

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

GEORGE MASON UNIVERSITY

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
  - …

GEORGE MASON UNIVERSITY

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

GEORGE MASON UNIVERSITY

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?

GEORGE MASON UNIVERSITY

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

GEORGE MASON UNIVERSITY