Ask An Expert

Ali Mosleh

Distinguished University Professor

Director, UCLA Garrick Institute for the Risk Sciences

ESREL 2018
Norwegian University of Science and Technology
Trondheim Norway

Experts and Expert Opinion

- The primary reason for using expert opinion is to deal with knowledge gap, uncertainty, in selected issues.

- An expert is an individual with specialized knowledge or skill in some specific domain.

- Expert Opinion can be viewed as expression of the judgment of an expert on a subject or issue. An opinion is usually regarded as impression, personal assessment, or subjective estimation of a quality or quantity of interest.

- Source for exploring unknown issues that are otherwise inaccessible

- Expert opinion, in contrast with factual information, is a judgment or a belief that, at least in the mind of the receiver of the opinion, is based on uncertain information or limited knowledge.
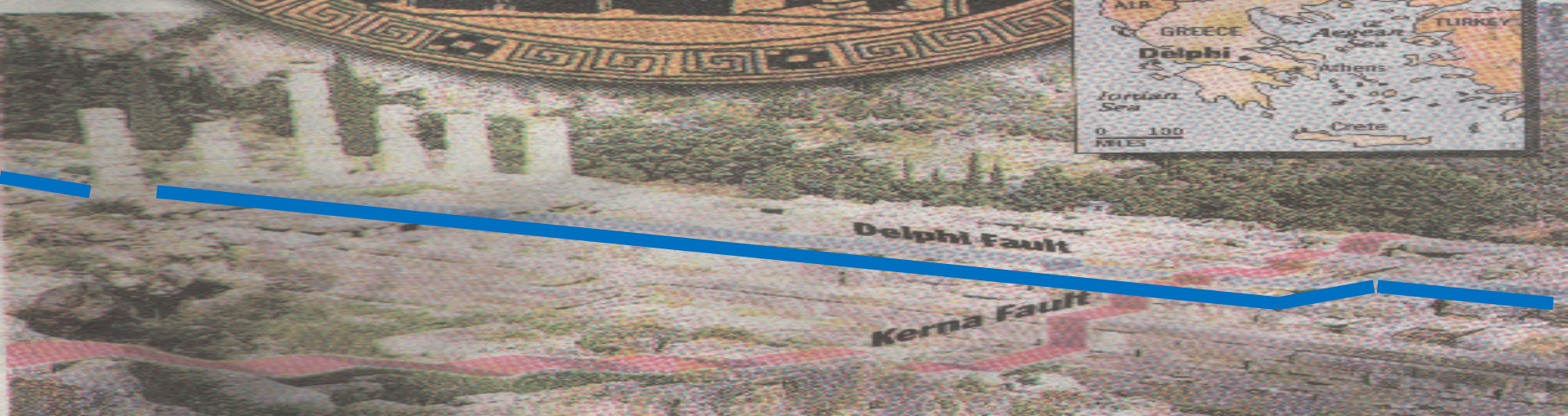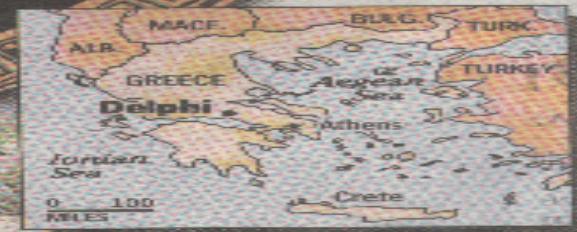
Expert Predictions of Crude Oil Price
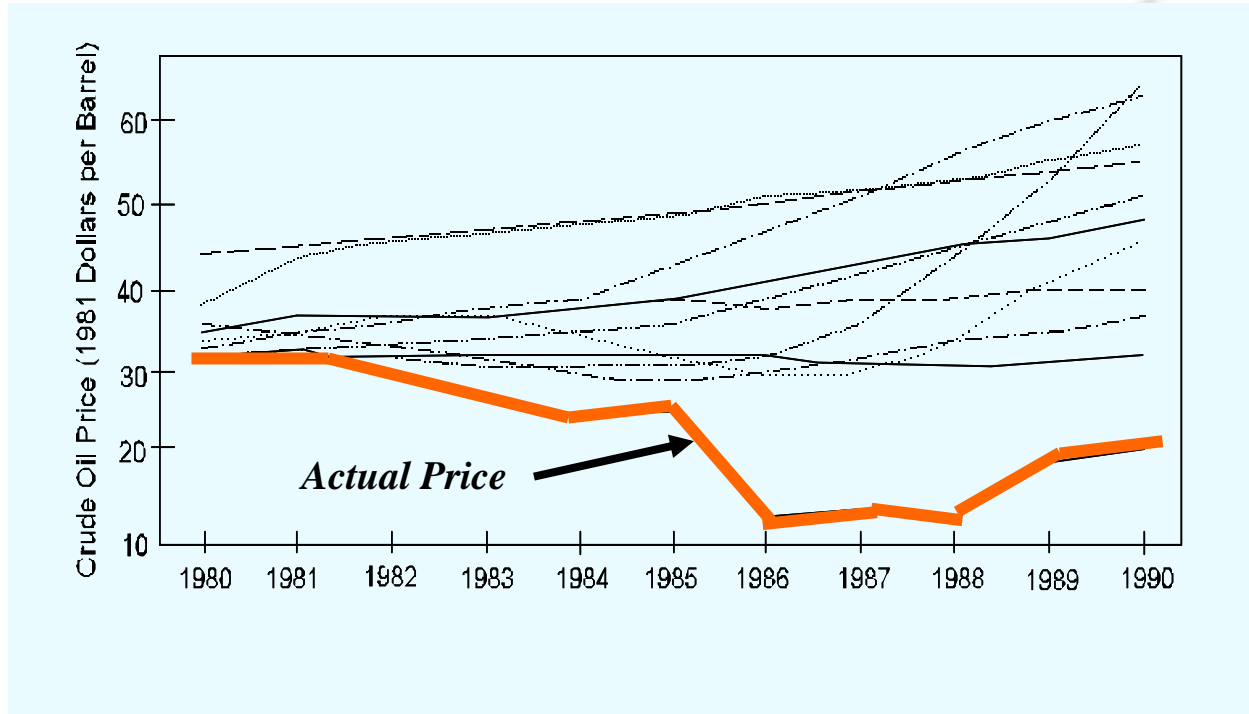
he *legendary oracle.*

A design from a Greek cup dated about 440 B.C. depicts an oracle seated on a tripod in the temple of Apollo.

MACE.   BULG.   TURK.

ALB.

GREECE   *Aegean*
                Sea   TURKEY
Delphi

                Athens

Ionian
Sea
                Crete
0   100
MILES

Delphi Fault

Kerna Fault

Measures of Quality of Expert Opinion

- **Substantive Goodness**: knowledge of the expert relative to the problem at hand

- **Normative Goodness**: expert's ability to express that knowledge in the form and metrics of interest
  -
  - in accordance with the calculus of probabilities and in close correspondence with his/her actual opinions

Use of Expert Opinion

- **Elicitation**
  - How to select
  - One expert or many (panel and composition)
  - How to elicit the opinion
- **Use**
  - How to use
    - a) expert information, and
    - b) information about the expert,
    to estimate the unknown quantity.

  - In case of multiple experts, how to aggregate the opinions.

Expert Elicitation…

Factors Affecting Quality of Results

- expert qualifications,

- estimation procedure,

- elicitation process,

- aggregation method,

- and available calibration information (e.g., historical knowledge on expert performance)

Example of Selection Criteria

(NRC, 1997 for Seismic Hazard Analysis)

- Strong relevant expertise through **academic training** professional accomplishment and experiences, and **peer-reviewed publications**;
- Familiarity and knowledge of various aspects related to the issues of interest;
- Willingness to **act as proponents** or impartial evaluators;
- Availability and willingness to **commit needed time** and effort;
- **Specific knowledge** and expertise of the issues of interest;
- Willingness to effectively participate in needed debates, to prepare for discussions, and provide needed evaluations and interpretations; and
- Strong communication skills, interpersonal skills, flexibility, **impartiality,** and **ability to generalize and simplify**

Expert Attributes vs. Performance
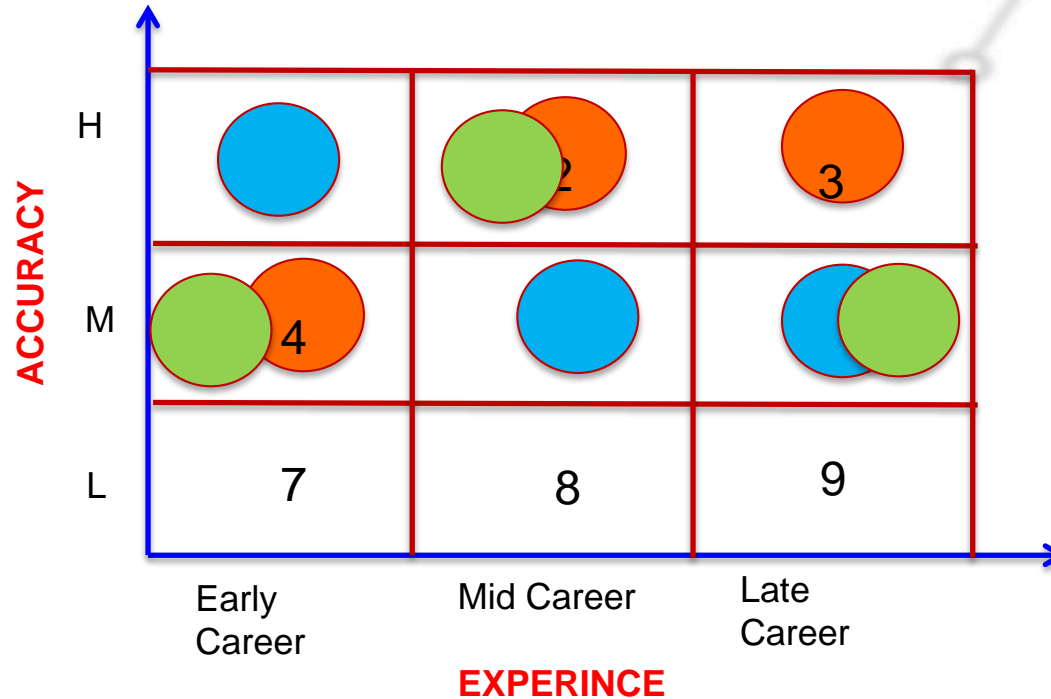
Expert Performance
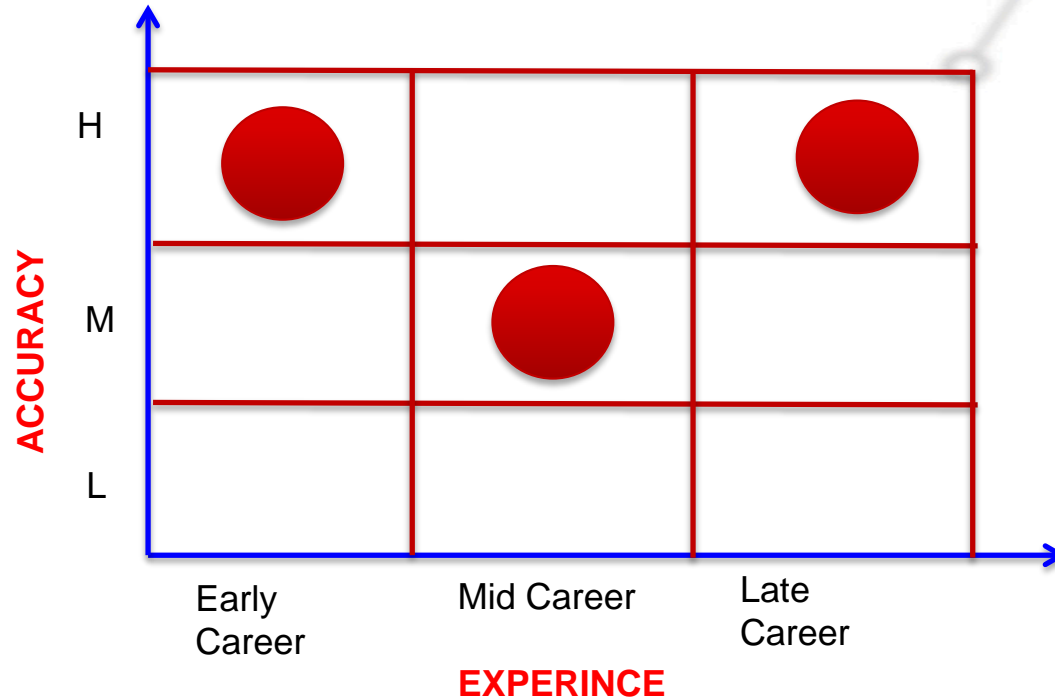
Expert Orientation

- Training on normative aspects
  - Experts might not be comfortable with the term "probability" but may answer about "events per year" or "recurrence interval."
  - Also notions such as "central tendency," and "dispersion measures,"
- Possible use of indirect elicitation should be explored with the experts.
  - Conditional probabilities
- Sources of bias (including overconfidence and base-rate fallacy) should be discussed.
- Include a search for any motivational bias of experts. These motivational biases, once identified, can sometimes be overcome by redefining the incentive structure for the experts.

Some Causes of Bias

<u>Availability</u> recalling events or situations similar to the event or issue of interest. Therefore, probabilities of well-publicized events tend to be overestimated, whereas probabilities of unglamorous events are underestimated.

<u>Anchoring</u> tendency to start with an initial estimate and correct it to the issue at hand. However, the correction might not be sufficient.

<u>Representativeness</u> tendency to evaluate intuitively the conditional probability P(BIA) by assessing the similarity between A and B. The problem: similarity is symmetric whereas conditional probabilities are not.
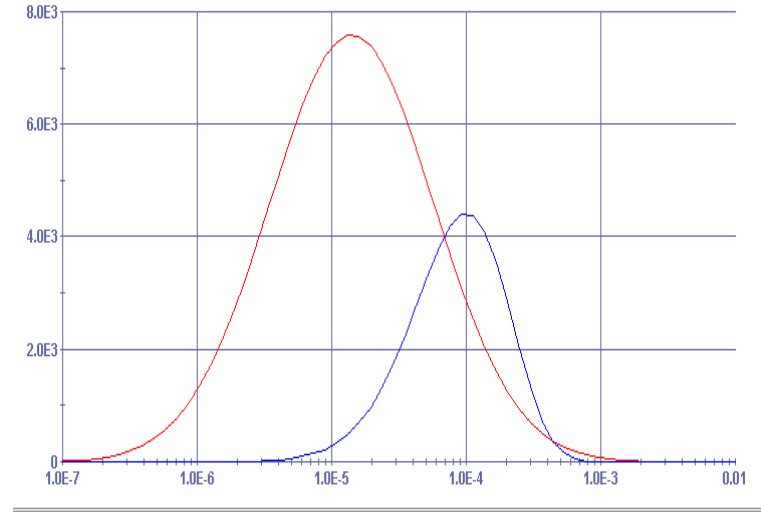
<u>Control factor</u> perception of subjects in that they can control or had control over outcomes related to an issue at hand. ⎯**Opinion Shapers**

Expert Calibration and Use …

Engine of Inference

$$Pr(A|E) = \frac{Pr(E|A)Pr(A)}{Pr(E)}$$

Expert Opinion in Bayesian Framework

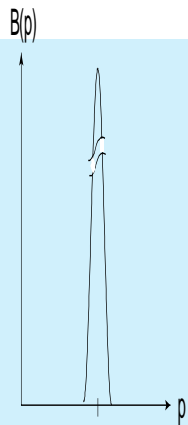$$\pi\left(x|x'\right) = \frac{L\left(x'|x\right)\pi_0\left(x\right)}{\int L\left(x'|x\right)\pi_0\left(x\right)dx}$$

X = Unknown to be estimated

X' = Expert estimate

Limiting Cases

- Perfect Expert
- $L(x^*|x) = \delta(x^*- x)$

$$\pi\big(x^*\big|x\big) \;=\; \frac{\delta\big(x^*-\,x\big)\,\pi_0(x)}{\displaystyle\int \delta\big(x^*-\,x\big)\,\pi_0(x)\,dx} \;=\; \delta\big(x^*-\,x^*\big)$$

- Non-Expert
- $L(x^*|x) = c(x^*)$

$$\pi\big(x\big|x^*\big) \;=\; \frac{c\big(x^*\big)\,\pi_0(x)}{\displaystyle\int c\big(x^*\big)\,\pi_0(x)\,dx} \;=\; \pi_0(x)$$

Additive Error Likelihood Model

$$X^* = x + E$$

$$\overline{X}_1^* = x + b_1$$

$$L\left(x_1^*|x\right) = \frac{1}{\sqrt{2\pi}\ \sigma_1}\ e^{-\frac{1}{2}\left[\frac{x_1^* - \left(x + b_1\right)}{\sigma_1}\right]^2}$$
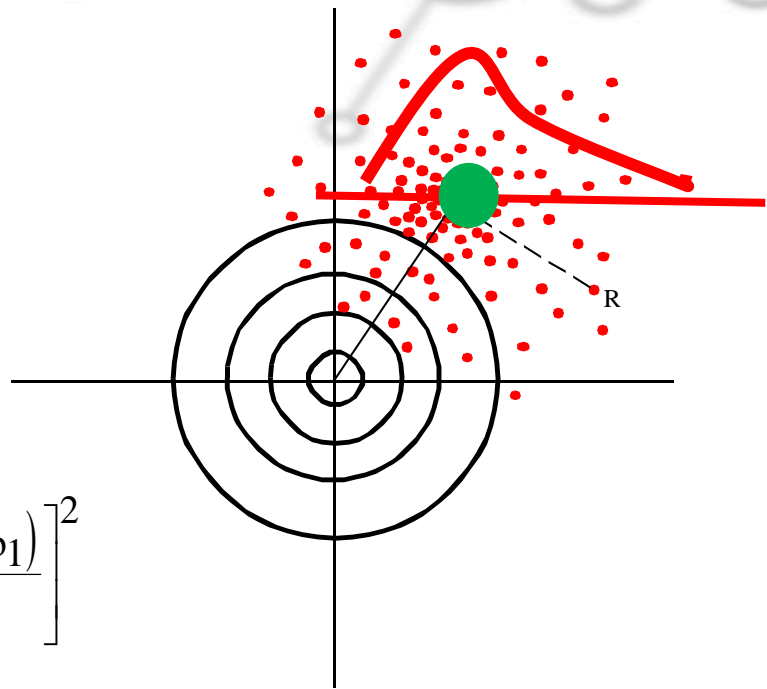
Additive Error Model Results

- Likelihood

$$L\left(x_1^*\middle|x\right) = \frac{1}{\sqrt{2\pi}\ \sigma_1}\ e^{-\frac{1}{2}\left[\frac{x_1^* - \left(x + b_1\right)}{\sigma_1}\right]^2}$$

- Posterior: Normal Distribution

$$\overline{X}_p = \omega_0\, x_0 + \omega_1\left(x_1^* - b_1\right) \qquad \sigma_p^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}}$$

$$\omega_i = \frac{\sigma_p^2}{\sigma_i^2}, \qquad i = 0,1$$

Multiplicative Error Model Of Likelihood

$$X_1^* = x E_1$$

$$\ln X_1^* = \ln x + \ln E_1$$

$$L\left(x*|x\right) = \frac{1}{\sqrt{2\pi}\ \sigma_1\ x_1^*}\ e^{-\frac{1}{2}\left[\frac{\ln x_1^* - \left(\ln x + \ln b_1\right)}{\sigma_1}\right]^2}$$
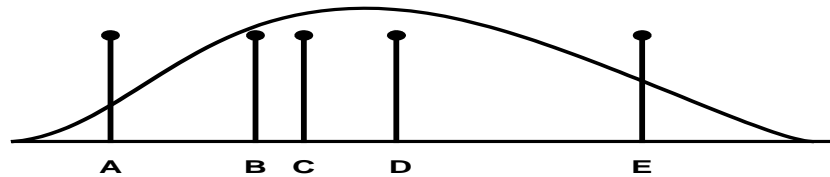
Aggregation

- The aggregation of multiple opinions tends to be more accurate than the opinion of a single expert.

- Approaches
  - mathematical
  - behavioral

Aggregating The Opinions Of Multiple Experts

- mathematical methods generally yield better results than "behavioral" / "consensus" methods, such as the **Delphi approach.**

- Issues in interactive group processes:
  - A "central tendency" effect.
  - The tendency for less confident members of the group to limit their participation.
  - Group pressures for conformity.
  - The strong influence of dominant personalities.
  - An investment in maintaining the integrity of the group itself.
  - A tendency to reach speedy decisions.

Many Experts

- The Reactor Safety Study (WASH-1400)

- Arithmetic Average

$$x_a = \sum_{i=1}^{N} \omega_i x_i^*$$

- Geometric Average (IEEE 500)

$$x_g = \prod_{i=1}^{N} x_i^{*\omega_i}$$

Aggregation– Bayesian Approach

$$\pi\left(x \middle| x_1^*, \mathsf{L}, X_N^*\right) = k^{-1} L\left(x_1^*, \mathsf{L}, X_N^* \middle| x\right) \pi_0(x)$$

$$L\left(x_1^*, \mathsf{L}, x_N^* \middle| x\right) = \prod_{i=1}^{N} L_i\left(x_i^* \middle| x\right)$$

Ad-Hoc methods such as WA or GA are shown to be special cases of the Bayesian method

Dalky's Impossibility Theorem (1974)

- Two Experts 1, 2
- Two Events A, B
- Expert Estimates: A_1, B_1, A_2, B_2

$A = 0.5*A_1 + 0.5*A_2$

$B = 0.5*B_1 + 0.5*B_2$

$C = A*B$

$= [0.5*A_1 + 0.5*A_2][0.5*B_1 + 0.5*B_2]$

$C1 = A_1*B_1$

$C2 = A_2*B_2$
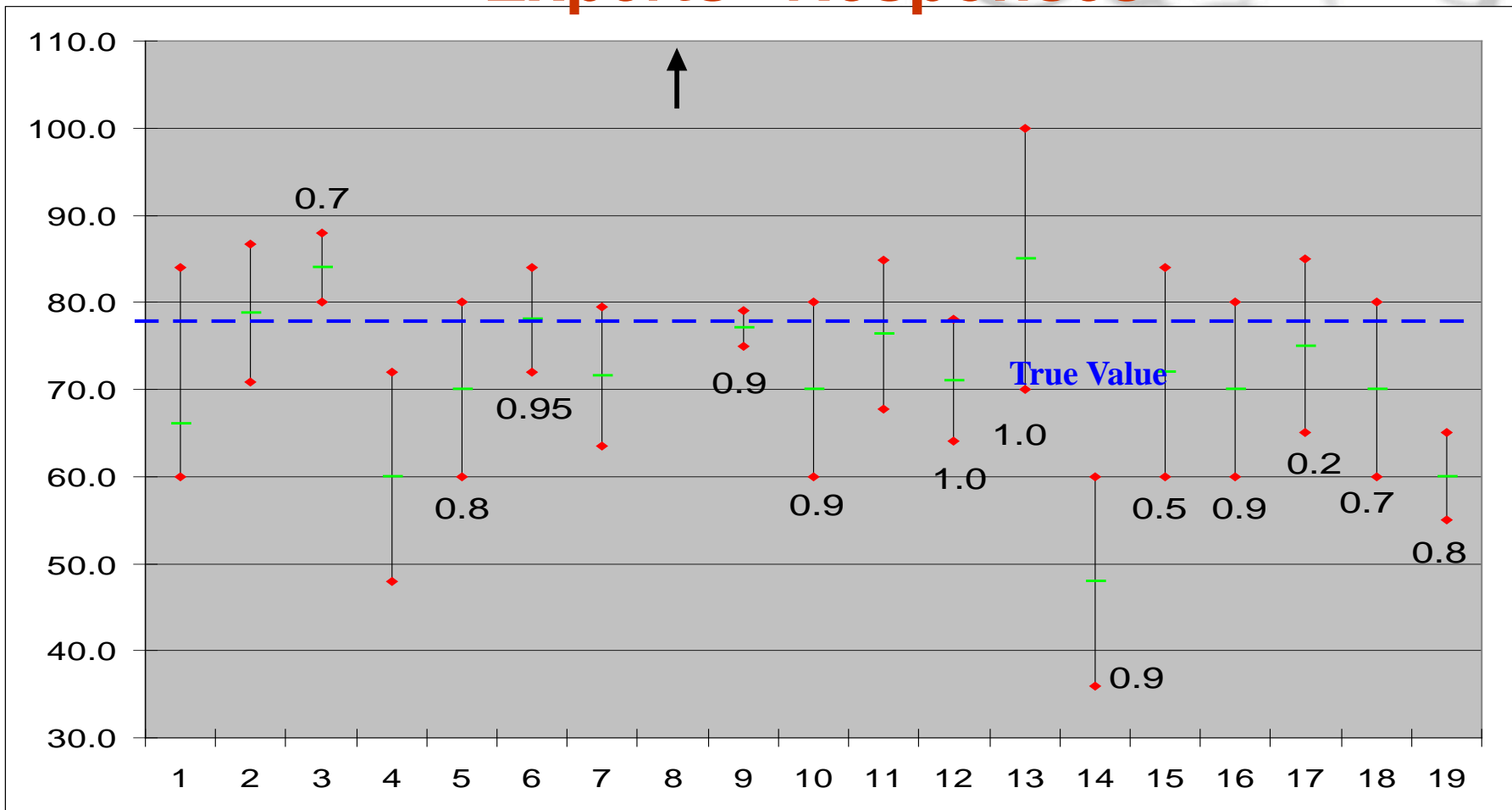
$C = 0.5*C_1 + 0.5*C_2$

$C = 0.5*A_1*B_1 + 0.5*A_2*B_2$

Keeping The Skeptics in the Room
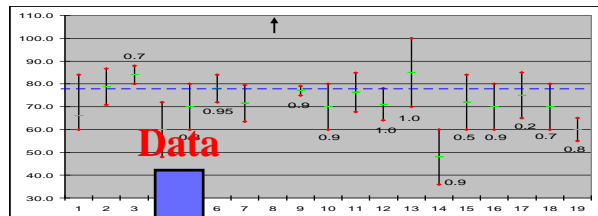
Experts' Responses

Experts' Responses

- 19 responded

- The true value was within the range of 10 responses, and outside of while 9 others

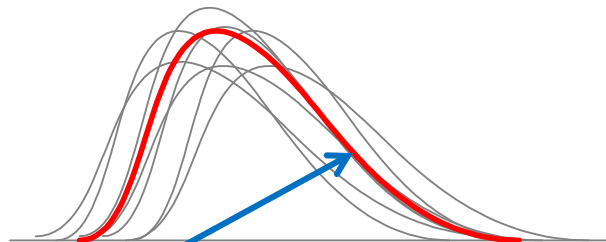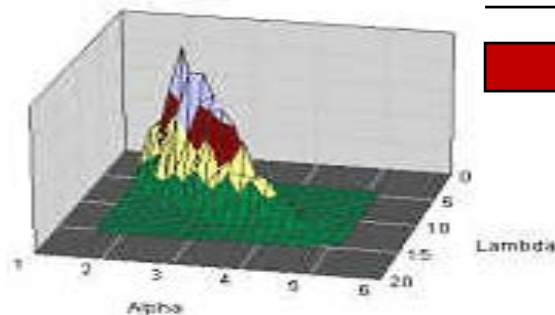- Average confidence level was 0.72, compared with observed 9/19 = 0.47, indicating overconfidence

Bayesian Aggregation Method

$$\pi(\mu,\ \sigma\,|\,E)\ =\ \frac{L(E\,|\,\mu,\ \sigma)\,\pi_o(\mu,\ \sigma)}{\int_\mu \int_\sigma L(E\,|\,\mu,\ \sigma)\,\pi_o(\mu,\ \sigma)\,d\mu\,d\sigma}$$

$$E = \{k_i,\ T_i\}\ i\ =\ 1,\ K,\ N$$

Bayes Theorem

$$\overline{\phi}(\lambda)\ =\ \iint_{\mu\ \sigma}\ \phi(\lambda\,|\,\mu,\sigma\,)\,\pi(\mu,\sigma\,|\,E)\,d\mu d\sigma$$

Result
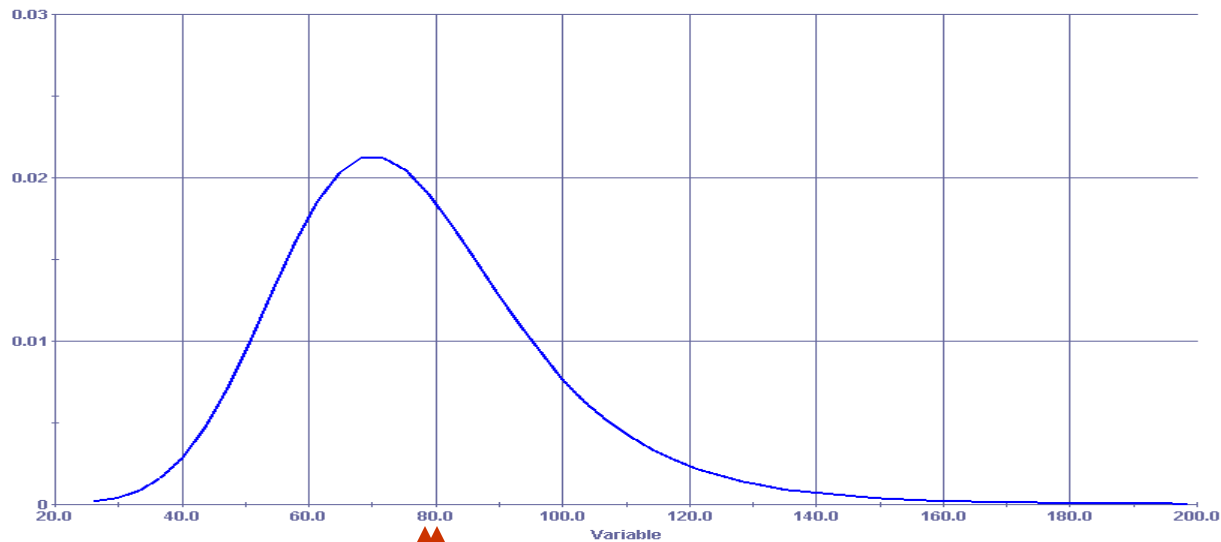
Mean of Posterior = 80 inches
True Value = 78 inches

Meta Data Analysis
# Expert Opinion Calibration and Aggregation

- The possibility of developing "generic" likelihood function based on large numbers of observed expert errors in different domains

- Whether use of such generic likelihoods would reduce future prediction errors

- Does mathematical aggregation provide a better estimate ?
  - What is the best method for aggregations ?
    - Various averaging techniques ?
    - Bayesian methods ?
    - Other ?

- How many experts?

- Does the type of quantity estimated, "physical"—variables having units of mass, time, etc.—or "probabilistic" matter?

Expert Judgment Data Source

Records/Predictions/Themes↓EJDS Source→	TUD	UMD	EJE Total	TD	EDJS Total
Number of records in Physical Category	540	1,181	1,721	11	1,732
Number of records in Probabilistic Category	66	1	67	26	93
TOTAL	*606*	*1,182*	*1,788*	*37*	*1,825*
Number of predictions in Physical Category	4,661	1,445	6,106	130	6,236
Number of predictions in Probabilistic Category	516	13	529	60	589
TOTAL	*5,177*	*1,458*	*6,635*	*190*	*6,825*
Number of themes in Physical Category	27	16	43	5	48
Number of themes in Probabilistic Category	8	1	9	5	14
TOTAL	35	17	52	10	62

Generic Calibration of Experts

Distribution of ratio of Elicited predictions, **e'** and realized values, **e**

Binned ln (e/e') mass is more concentrated for physical data than for probabilistic data
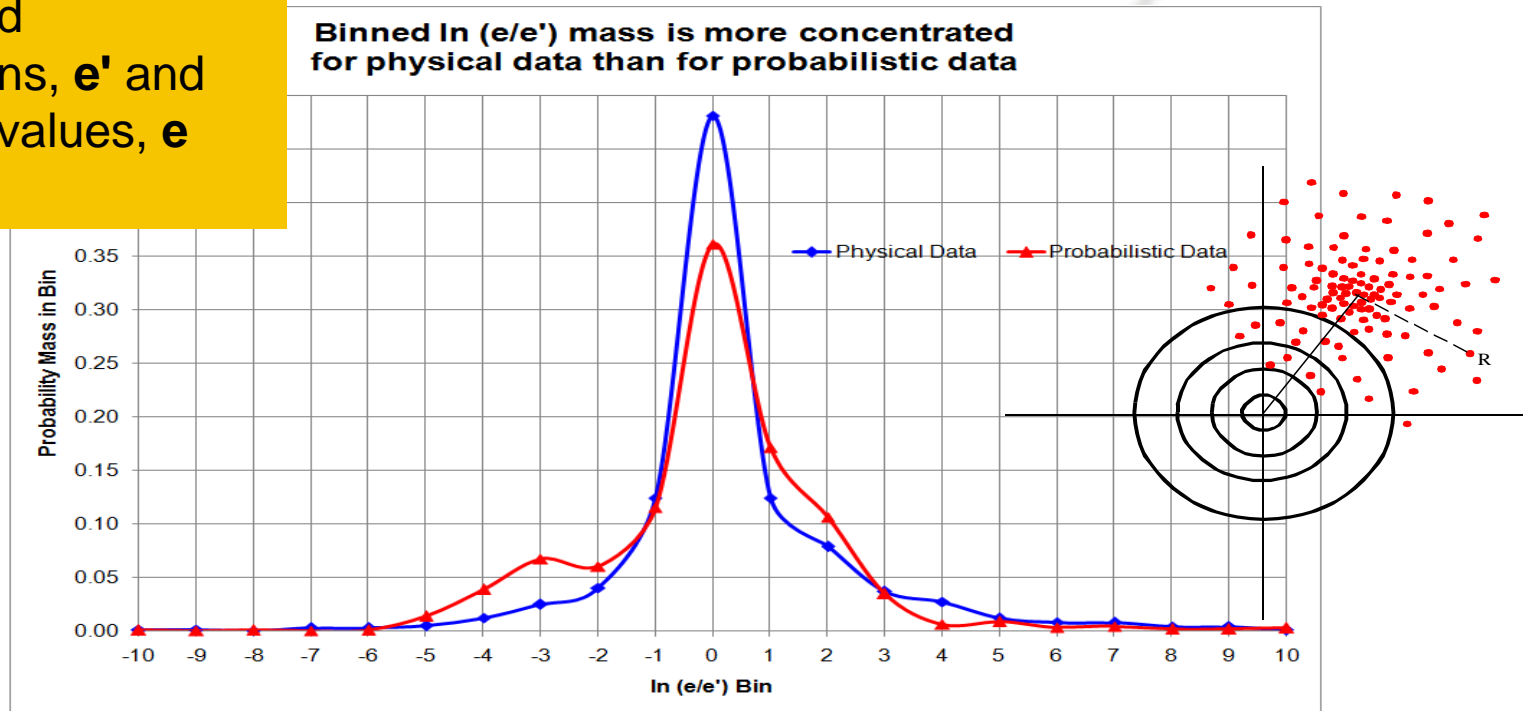
Evidence On Accuracy Of Expert Opinion In Engineering Risk Assessment

**Comparison of data and expert opinion on the distribution of component maintenance time**

| Component Type	Technical Specification Time*	Characteristics of Distribution				Observed/ Predicated	
		Data-Based		Expert-Estimated			
		Range Factor	Mean	Range Factor	Mean	For Range Factor	For Mean
Pumps	None	22.1	265	6.2	116.0	3.56	2.28
	168 Hours	6.2	29	1.8	40.4	3.44	0.72
	72 Hours	5.9	11	1.5	20.9	3.93	0.53
	# 24 Hours	4.2	7	1.5	10.8	2.80	0.65
Valves	None	26.2	135	6.2	116.0	4.23	1.16
	72 or 168 Hours	5.2	19	1.8	40.4	2.89	0.47
	# 24 Hours	3.8	4	1.5	20.9	2.53	0.19
Heat Exchanges	None	4.6	580	6.2	116.0	0.74	5.03
Other**	None	11.0	39	6.2	116.0	1.77	0.34
	> 72 Hours	3.0	37	1.8	40.4	1.67	0.92
	48 or 72 Hours	7.3	14	1.5	20.9	4.87	0.67
	# 24 Hours	5.8	6	1.5	10.8	3.87	0.56

* Limit of allowable downtime
** For example, diesel generators, fans, electrical equipment; also includes heat exchanges with technical specifications.
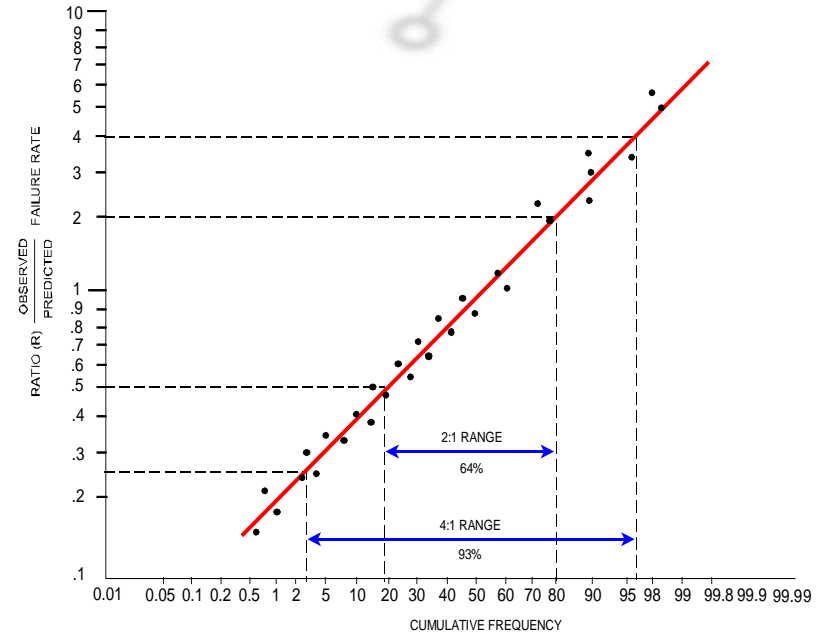
Overestimation Probability by Factor
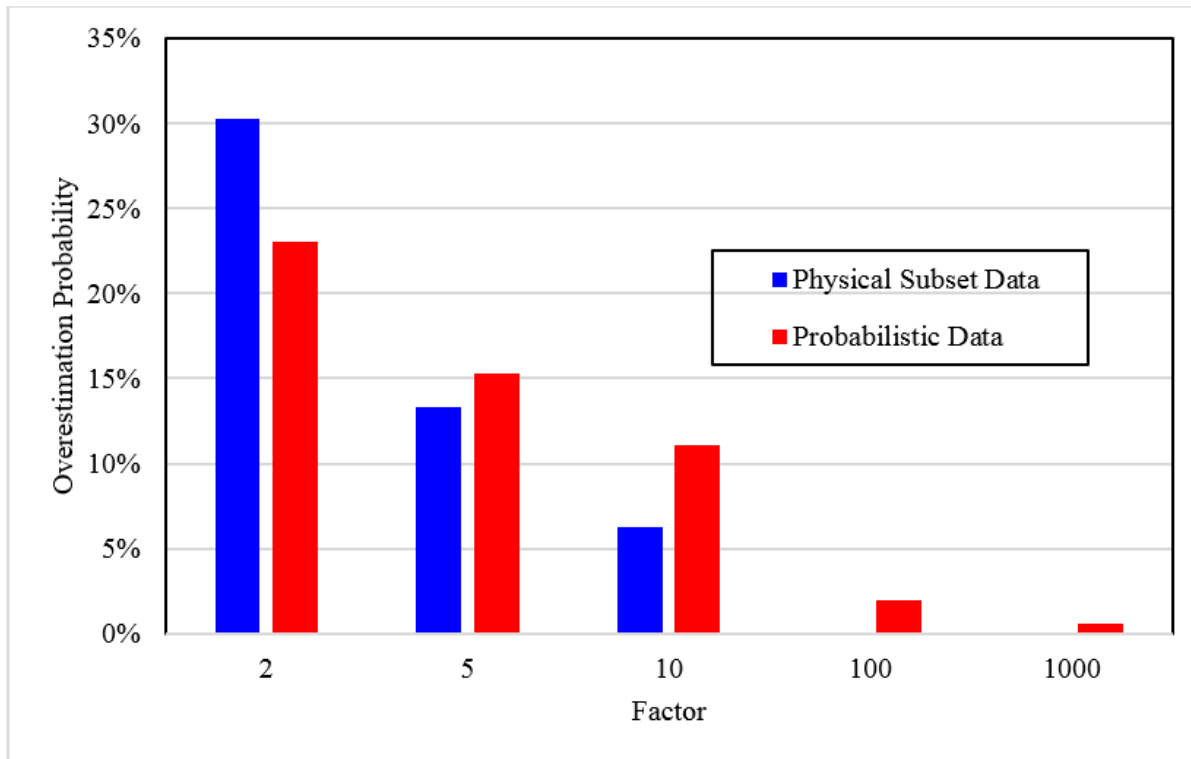
# Underestimation Probability by Factor
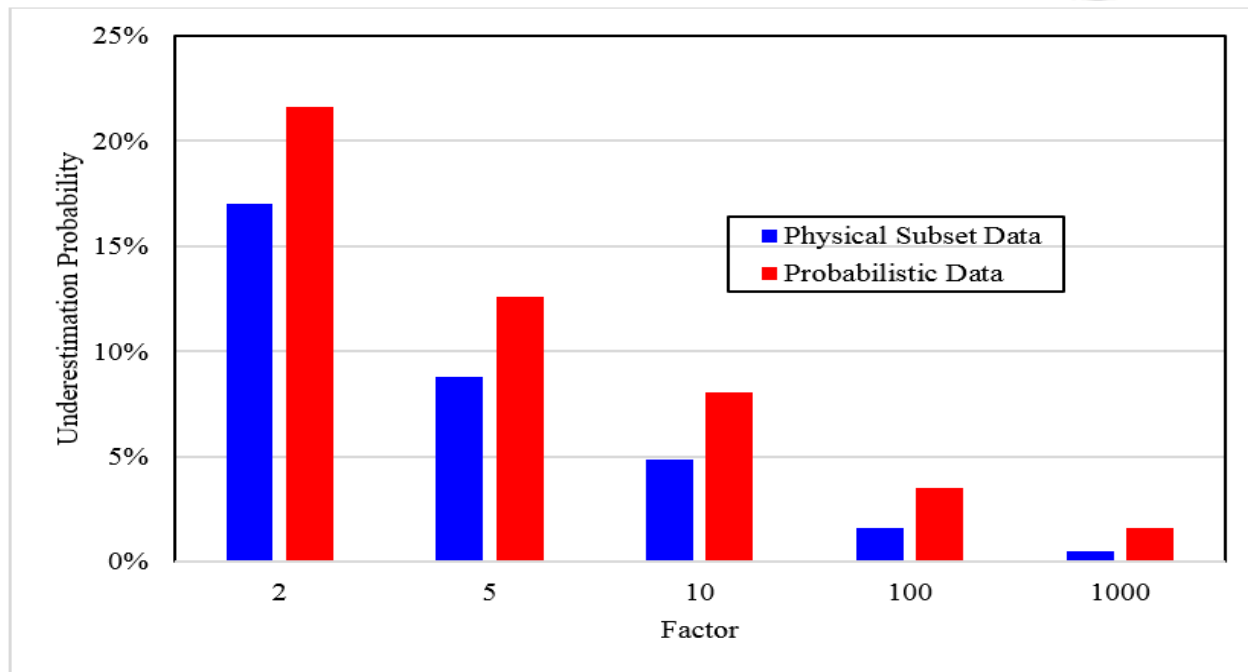
Size and Composition of Expert Panel

- "N-heads rule" suggests under certain assumptions, the higher the number of experts the better the result

- If possible the panel should be large enough to capture complementary expertise and achieve diversity of opinion, to ensure a balanced and broad spectrum of viewpoints, expertise, and technical points of view
  - Experts familiar with the specific technical subject
  - Experts in the broader domain of knowledge
  - Experts in support areas, and related domains, such as statistics, risk analysis, and decision-making
  - Observers, discussion facilitators, and expert opinion elicitation experts

N-Heads Effect – Physical Quantity Estimates

Weighted Average MME vs Number of Heads, s for Physical Data
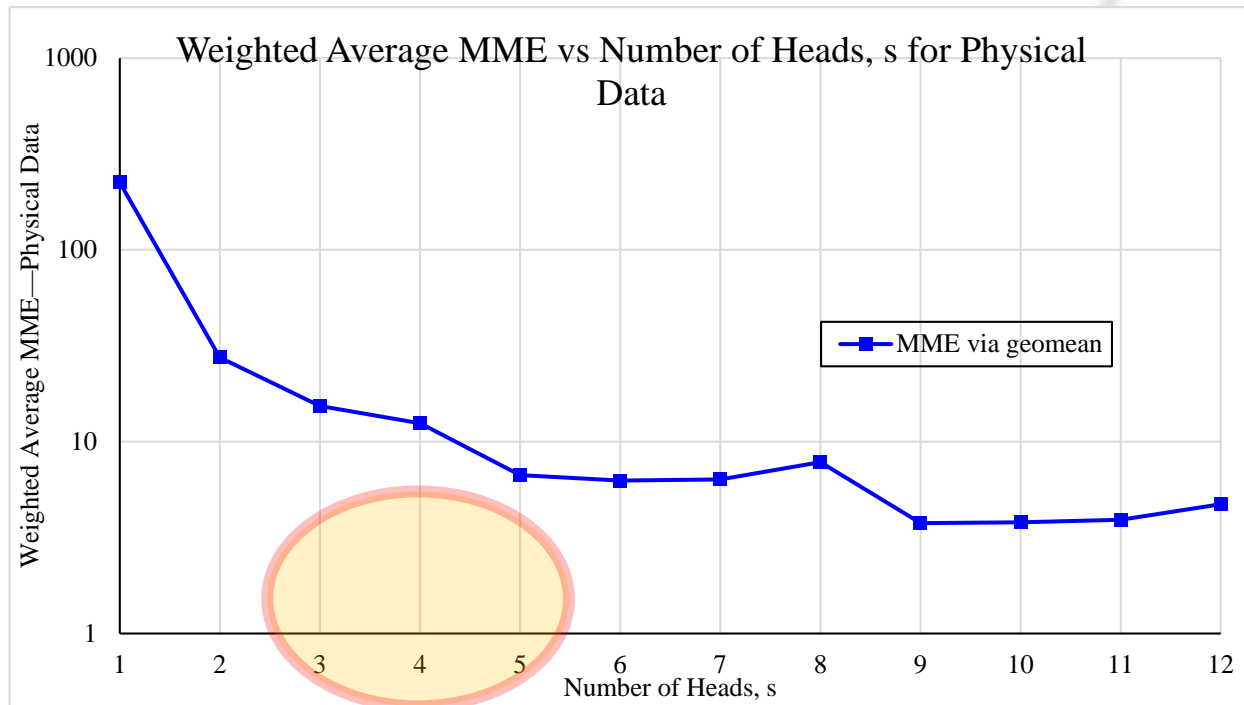
N-Heads Effect – Probability Estimates
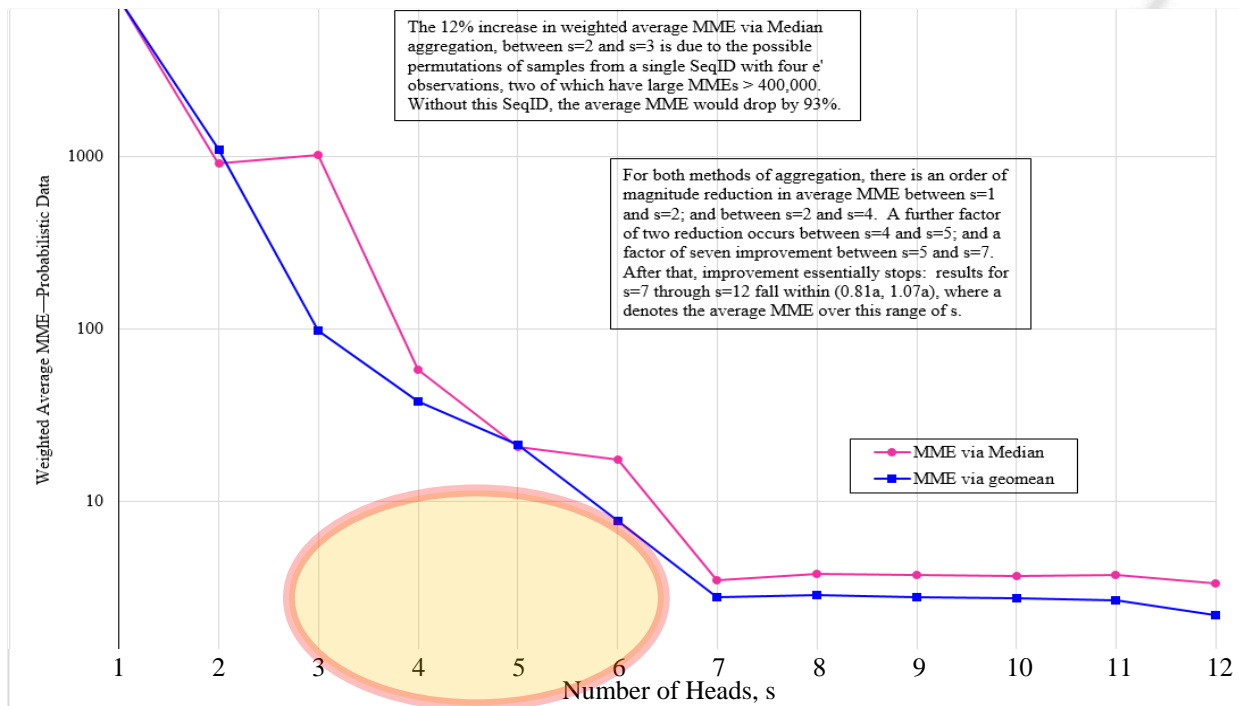
Summary of Some Helpful Findings

- Select good domain experts, train them on normative aspects
- Aggregation of opinion of multiple experts tends to give more accurate results than the opinion of a single expert (N-Heads Rule)
- Mathematical methods of aggregation are generally preferable to behavioral methods for reaching consensus.
- Quality of judgments can be substantially improved by decomposing the problem into a number of more elementary problems
- There is a significant improvement in the overall results if the initial problem definition and decomposition is done with care and in consultation with the experts

Summary of Some Helpful Findings (cont.)

- **Expert opinion is subject to biases**
  - The possibility of systematic overestimation or underestimation
  - Overconfidence; i.e., the tendency for people to give overly narrow confidence intervals which reflect more certainty than is justified by their knowledge about the assessed quantities
- **Effective techniques to reduce overconfidence:**
  - use of calibration techniques, and
  - encouraging experts to actively identify evidence that tends to contradict their initial opinions.
- **Sources of strong dependencies among experts should be identified**
  - Weak dependence does not seem to have a major impact on the value of multiple expert judgment.

Rich Body of Literature

- Increasing sophistication of elicitation methods
    - Selection, attributes
    - Elicitation process
- Progress in generic calibration
    - Domain-specific
- Studies on
    - Performance and effectiveness of aggregation methods
    - Understanding and dealing with sources of dependencies