



The Unseen: Statistical Inference with Limited Data

Trevor VanAtta, US Army TACOM
Joint SCEA/ISPA Conference
Orlando, FL (June 2012)



Agenda

- ▶ Calculation or Intuition? *Probability vs. Confidence*
- ▶ Heuristics and Biases
 - Representativeness
 - Availability
 - Anchoring and Adjustment
 - Reducing Errors due to Heuristics
- ▶ Formulating Distributions with Limited Data
- ▶ The Unseen: Informing Decision Makers with Cost Risk Analysis



Calculation or Intuition?

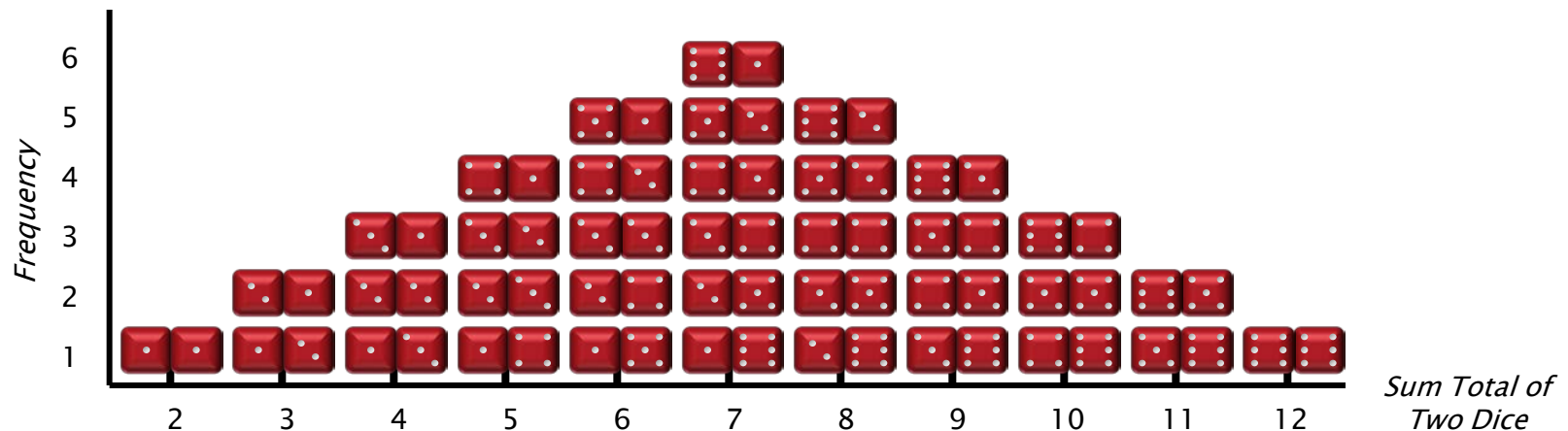
Probability vs. Confidence



Calculating Probability

Probabilistic judgments can be made with questions that are:

- 1) Decompositional: All possible outcomes can be determined
- 2) Frequentistic: Infinitely repeatable
- 3) Algorithmic: Results are measurable



“Objective measurements of probability are often unavailable, and most significant choices under risk require an intuitive evaluation of probability.”

-Daniel Kahneman & Amos Tversky



Calculation or Intuition?

- ▶ When forecasting cost, a complete *decomposition* of all future possibilities is unachievable
- ▶ Cost estimates require a combination of calculation and intuition to formulate *levels of confidence*
- ▶ Confidence = *Perceived Probability*

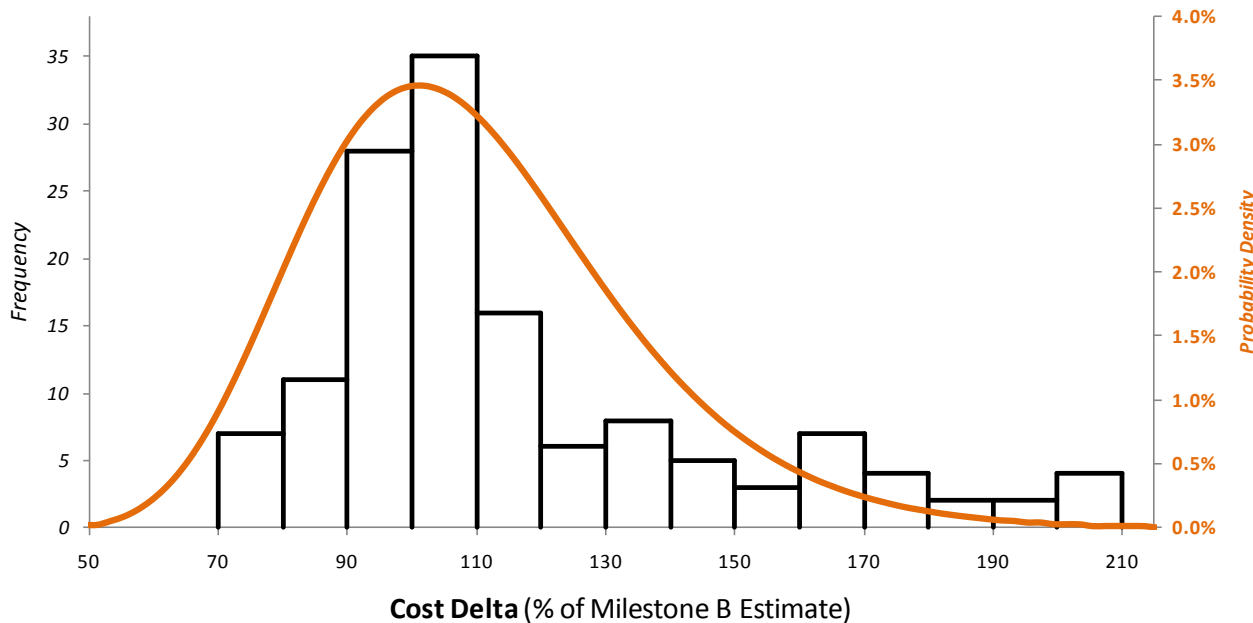
Scenario	Decompositional?	Frequentistic?	Algorithmic?
The Sum Total of Two Dice	Yes	Yes	Yes
Forecasting Future Costs	No	No	Yes
Evaluating Job Applicant	No	No	No

“A good forecast is a compromise between a point estimate, which is sure to be wrong, and 99.9% confidence interval, which is often too broad. The selection of hypotheses in science is subject to the same trade-off. A hypothesis must risk refutation to be valuable, but its value declines if refutation is nearly certain. *Good hypotheses balance informativeness against probable truth.*” –Daniel Kahneman



Are Point Estimates Trustworthy?

Distribution of the *Mistakes* Component of Procurement Cost Growth from Milestone B Baseline for 138 MDAPS (1970 - 1997)



“Mistakes” includes:

- Inadvertent Error
- Unstable Budget
- Ineffective Management

Corrected for:

- Quantity Changes
- Major Requirements Shifts

Lognormal Distribution
Mode = 100%; CV = 25%

Data Source: McNicol, David J.; “Cost Growth in Major Weapon Procurement Programs”, Second Edition; Institute for Defense Analysis (IDA); 2005



Can't we just make conservative assumptions?

NATO Analysis in Early 1950s of US and Soviet End-Strength



VS.

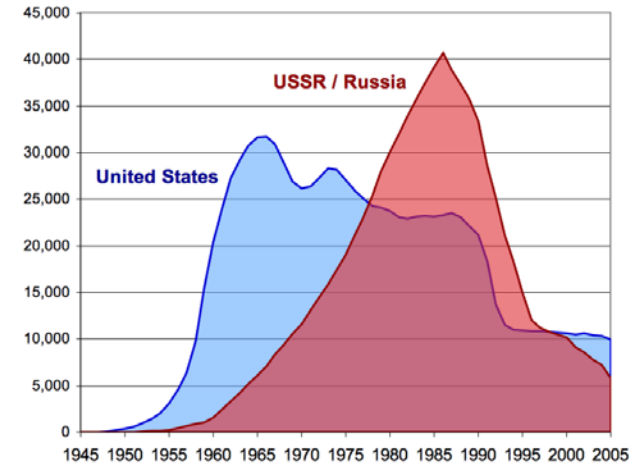


25 ill-equipped, ill-trained divisions unready for combat.

175 well-equipped, well-trained divisions ready for combat.

Resulting Conclusion: NATO can't compete with the Warsaw Pact's conventional ground and air forces, and must focus on nuclear deterrent.

U.S. and Soviet Nuclear Stockpiles, 1945-2005



However, later analysis refuted initial assumptions of 1950s analysis. In reality...

- 1) Fighting Power of one U.S. division was roughly equal to three Soviet divisions
- 2) "At least half of [the 175 divisions] were cadre divisions (that is, essentially paper units)"

*Sources:

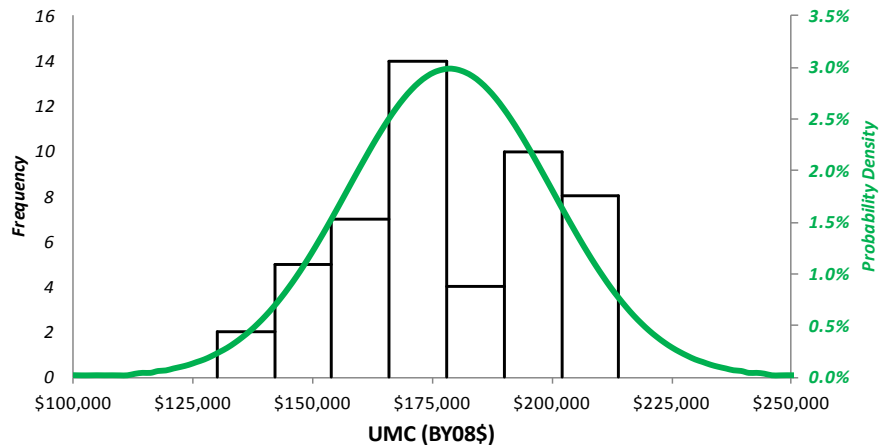
- 1) Enthoven, and Smith; *How Much is Enough? Shaping the Defense Program 1961-1968*; Chapter 4; 1971
- 2) Kristensen, Hans M. and Norris, Robert S.; "Global nuclear stockpiles, 1945-2006," *Bulletin of the Atomic Scientists*, Volume 62, No. 4 (July/August 2006), 64-66

Distributions from Data: CSDR Example

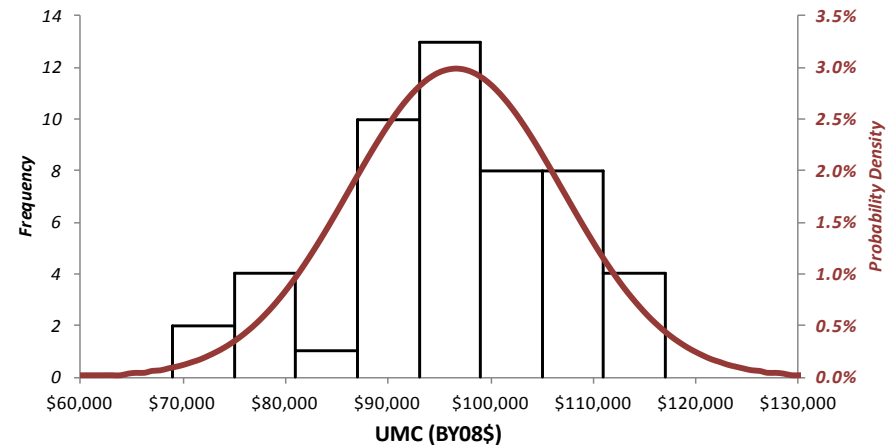


- ▶ Regression analysis and goodness-of-fit tests allow analysts to make probabilistic statements when ample data is available
- ▶ *Example:* I am 50% confident that Subsystem B will not cost more than \$100K

Unit Manufacturing Cost
Subsystem A



Unit Manufacturing Cost
Subsystem B

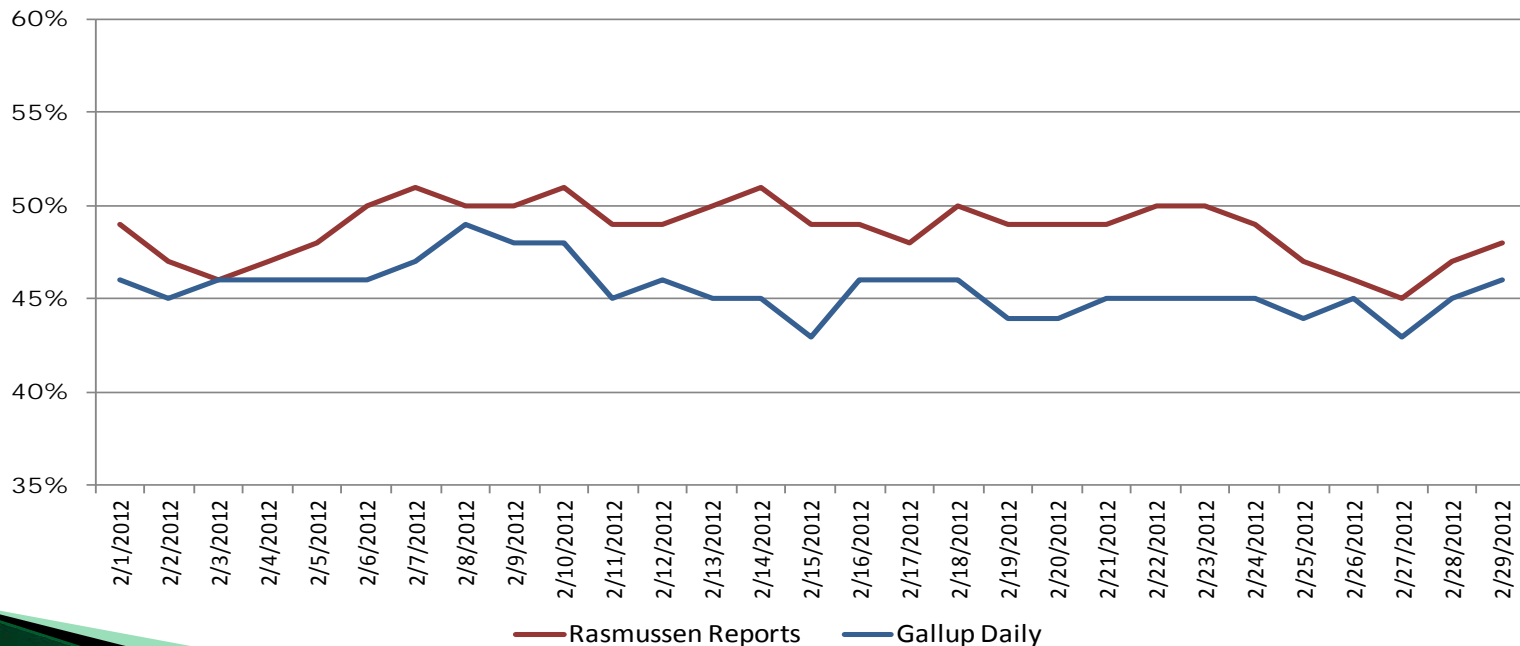




When Numbers Aren't Enough

- ▶ Gallup & Rasmussen both poll 1500 voters
- ▶ Both polls claim $\pm 3\%$ margin of error

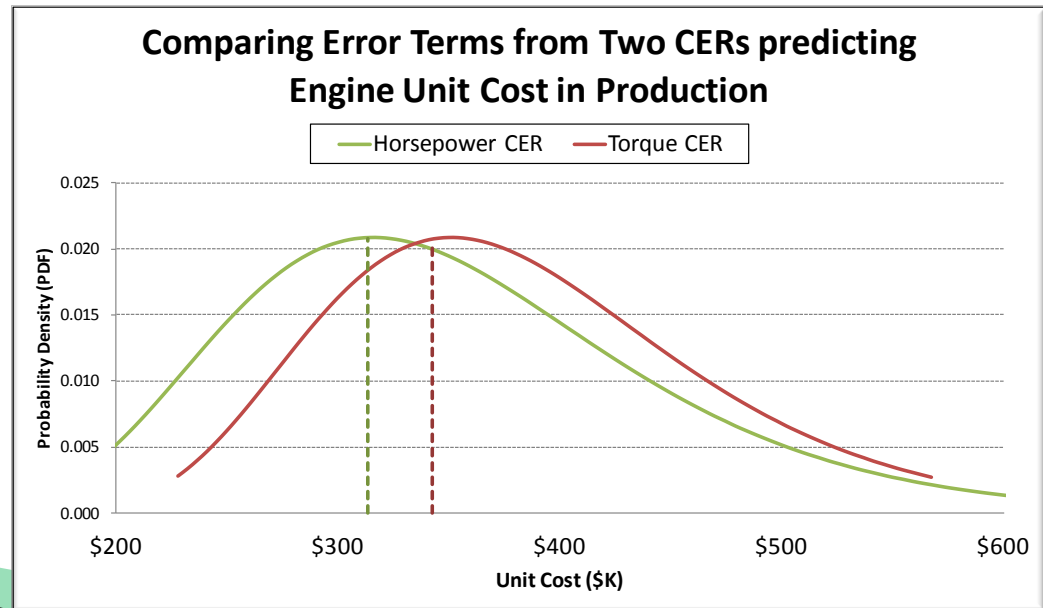
Presidential Approval Ratings
Gallup vs. Rasmussen (February 2012)





Confidence from CER Error Terms

- ▶ Scenario: Use data from existing engines to formulate CERs to predict the unit cost of other engines
- ▶ Issue: Two CERs resulted in different point estimates with different error terms. Which estimate should you choose?





The Collision of Intuition & Statistics: *Heuristics & Biases*



Heuristics and Biases

- ▶ ***What are heuristics?*** Mental shortcuts used to answer difficult questions when the answer is uncertain or unknown.

“When confronted with a difficult question, people often answer an easier one instead, usually without being aware of the substitution.”

–Daniel Kahneman, Princeton University, 2002 Nobel Laureate in Economics*

- ▶ ***Types of Heuristics:***
 - ▶ Representativeness
 - ▶ Availability
 - ▶ Anchoring & Adjustment

Representativeness: Linda and the Conjunction Rule



Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Based on the description above, which of the following statements about Linda is more probable?

- a) Linda is a bank teller.
- b) Linda is a bank teller and is active in the feminist movement.

RESULTS: 85% of respondents chose answer (b) in a clear violation of the conjunction rule.

Source: Tversky, Amos and Kahneman, Daniel; "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment"; *Psychological Review*, Volume 90, Number 4, 293-315 (October 1983)



Representativeness

- ▶ **Representativeness:** Judging the likelihood of a possibility based on how well it corresponds to a representative story.
- ▶ We often draw large inferences based on what we perceive to be a “representative” story or sample. *This causes us to favor narrow distributions over wider distributions.*

Example: What amount of money was spent on education by the U.S. federal government in 1987? Choose the answer that represents the best estimate:

- \$18 to \$20 billion
- \$20 to \$40 billion

After being told that the correct answer was \$22.5B, 80% of the study's participants answered “a) \$18 to \$20 billion” rather than the wider, more accurate answer.

Source: Foster, Dean P. and Yaniv, Ilan; “Graininess of Judgment: An Accuracy–Informativeness Tradeoff”; *Journal of Experimental Psychology: General*, 21, 1509–1521 (1990)



Availability

- ▶ **Availability:** Judging the likelihood of a possibility based on how easily instances and associations can be pictured or constructed.
- ▶ When formulating estimates, this causes us to reach for whatever data we remember or know best, causing us to narrow our distributions, discounting potentially valuable information across the full range of possibility

Word Form	Average # of Words Listed in 60 Seconds	Estimated Appearances in a 2000 Word, Four Page Novel
____ing	6.4	13.4
_____n_	2.9	4.7

Source: Tversky, Amos and Kahneman, Daniel; "Availability: A heuristic for judging frequency and probability"; *Cognitive Psychology, Volume 5, 207-232 (1973)*



Anchoring & Adjustment

- ▶ **Anchoring & Adjustment:** Estimating an uncertain value by clinging to a prominent reference point that we know to be wrong, then adjusting to a more likely value.
- ▶ We often don't adjust far enough because we tend to focus on information that lends credibility to the anchor value.
- ▶ Examples: In what year was George Washington elected President? (Anchor: 1776)

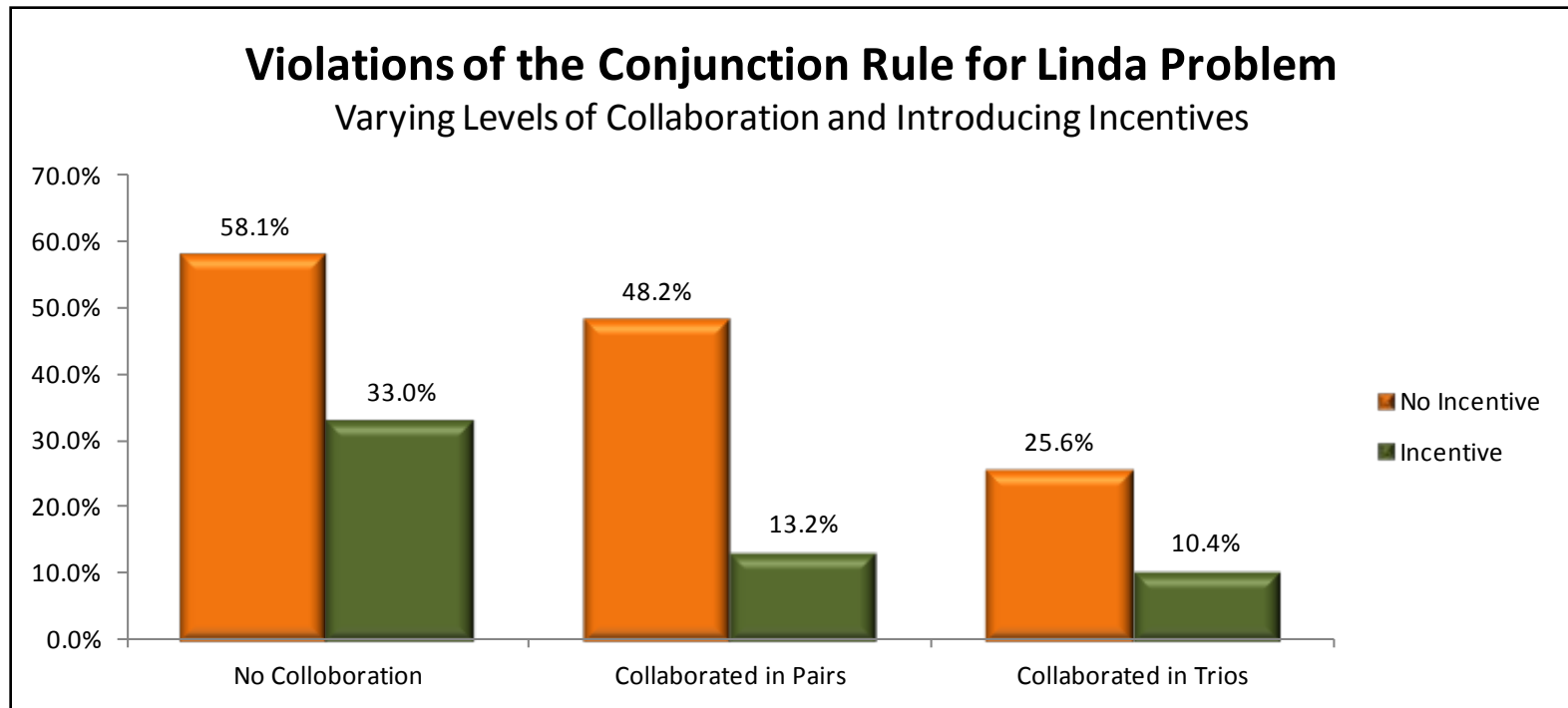
University of Arizona Study (1987) <i>Participants viewed a Tucson, AZ Home for 20 Minutes</i>				
List Price	\$65.9K	\$71.9K	\$77.9K	\$83.9K
Estimated Reasonable Price	\$63.6K	\$67.6K	\$70.1K	\$69.5K

Source: Neale, Margaret A. and Northcraft, Gregory B.; "Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions"; University of Arizona; Published in *Journal of Organizational Behavior and Human Decision Processes* 39, 84-97 (1987)



Reducing Heuristics

1) Encourage collaboration & incentivize accuracy



Source: Charness, Gary et al.; "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda"; 19 May 2009

Note: As an incentive, subjects were offered \$4.00 (US) for the correct answer



Reducing Heuristics

2) Encourage reasoned and objective skepticism*

When was Washington elected President?	
Anchor Value	1776
Nodded Head	1777.6
Remained Still	1779.1
Shook Head	1788.1

- ## 2) Increase the study of logic and statistics
- Statisticians score better in these types of studies
 - Subjects who practiced logic questions or were asked to “think as statisticians” also scored better

*Source: Epley, Nicolas and Gilovich, Thomas; "Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors"; *Psychological Science*, Volume 12, Number 5, 391-396 (September 2001)

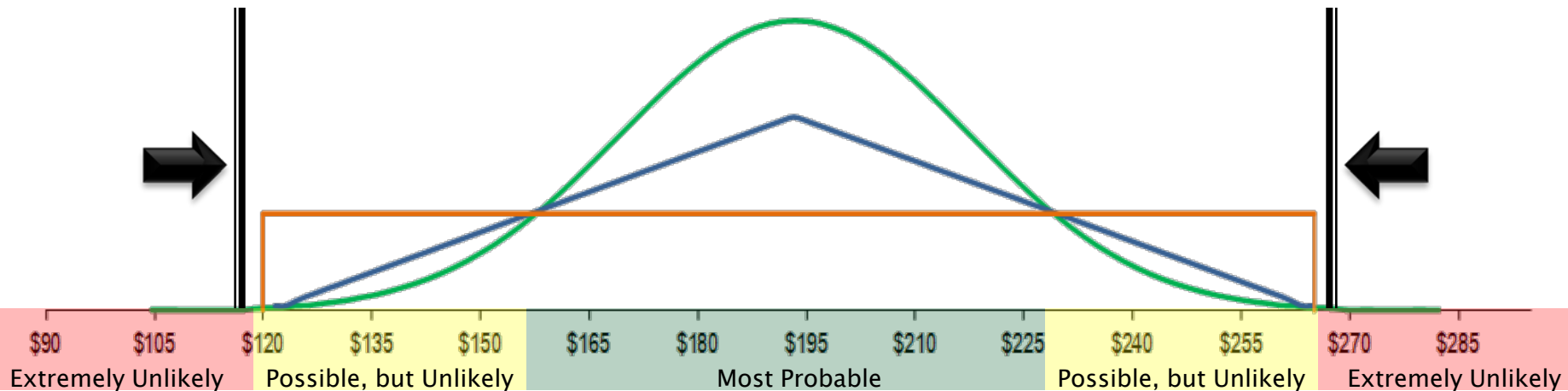


Formulating Distributions with Limited Data



Infinity Cropping

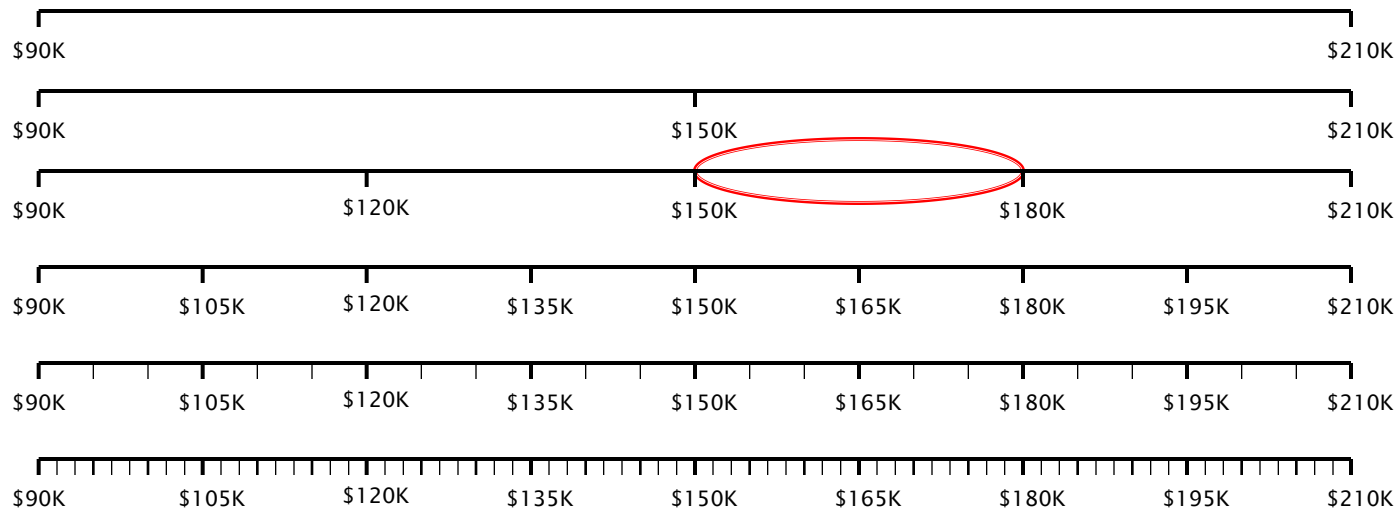
- ▶ Begin by using extreme values to “crop” out unlikely scenarios rather than searching for target value
- ▶ Begin with uniform distribution
- ▶ Narrow to Triangular, Pert, Beta, or Normal as more information discounts endpoints and uncovers more likely portions within the range





Grain Scales / Fidelity Intervals

- Subjects provided accurate range estimates 55% of the time using *grain scales* when asked to provide answers as if talking to a close friend
- In contrast, subjects asked to provide a 95% confidence intervals were only correct 43% of the time!



Hit Rate (%)	
Scale	Observed
1st	100%
2nd	51%
3rd	37%
4th	46%
5th	55%
6th	56%

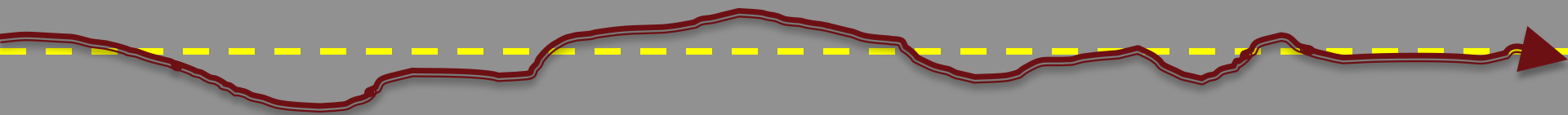
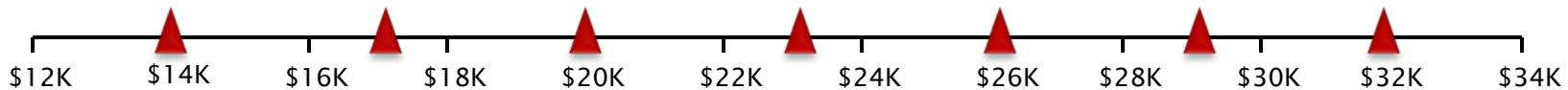
Source: Foster, Dean P. and Yaniv, Ilan; "Precision and Accuracy of Judgmental Estimation"; *Journal of Behavioral Decision Making*, Vol. 10, 21-32 (1997)



Eliminating the Average

- ▶ Averaging can lead to throwing away reams of informative data
- ▶ Averages act as anchor values, causing discounting of endpoints

Production Lot #	Reported Unit Cost of Engine
1	\$29,000
2	\$23,000
3	\$26,000
4	\$14,000
5	\$20,000
6	\$32,000
7	\$17,000

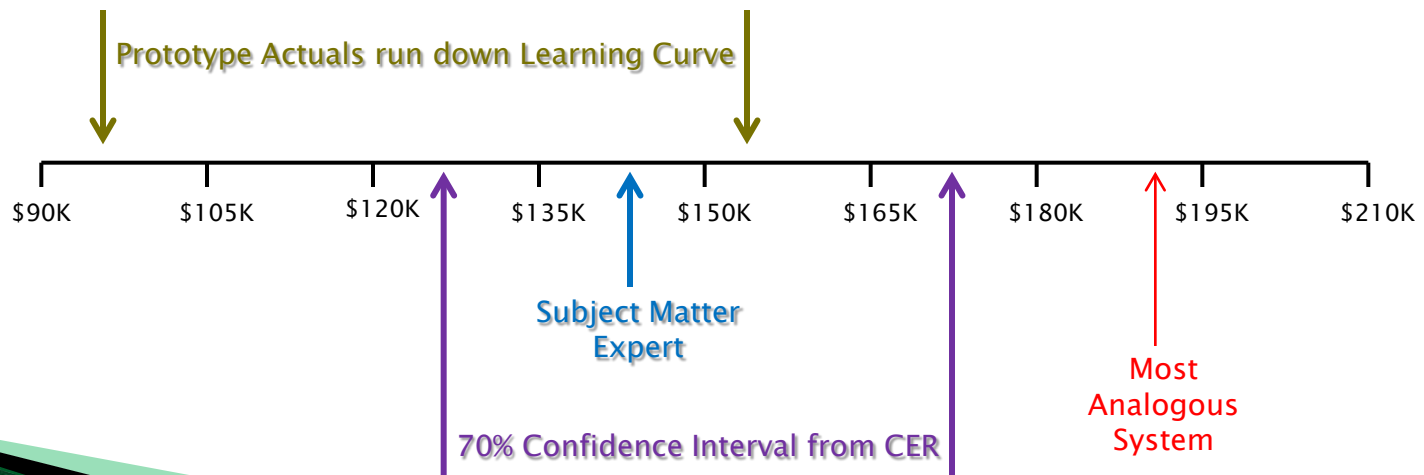


See: Savage, Sam; [The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty](#); Published by Wiley (2009); ISBN 978-1118073759



Multiple Methodologies

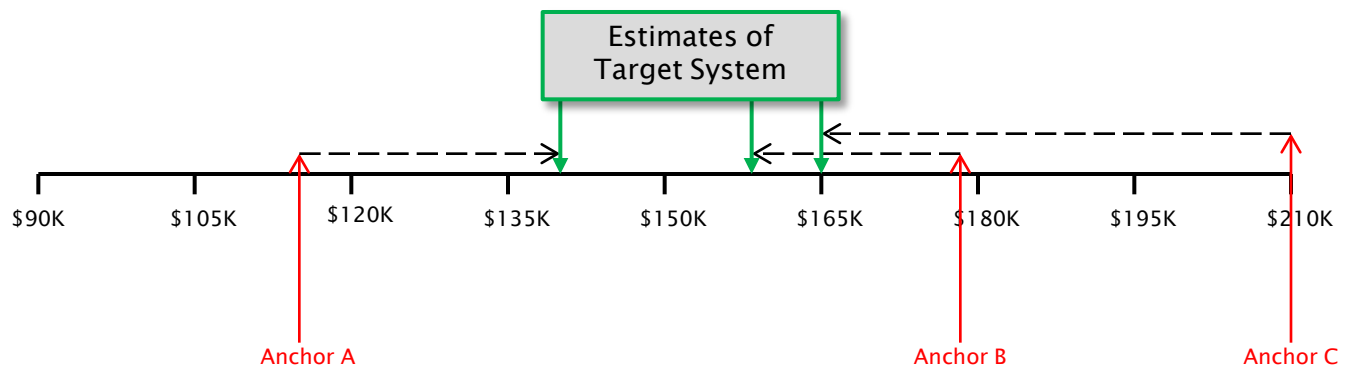
- ▶ **Convergence Tests** ask subjects to *converge* onto a single correct answer (Example: IQ Tests)
- ▶ **Divergence Tests** measure creativity via open-ended questions (Example: In 60 seconds, list as many uses for a brick as possible.)
- ▶ Cost risk analysis is a process of *divergence*.





Forced Anchoring

- ▶ Potential anchors can be split up into divergent estimating process
- ▶ Forcing multiple experts to provide estimates using different datasets can reveal more likely points within a wider range
- ▶ Helps convert Uniform Dist to Triangular, Pert, Normal, Beta, etc.

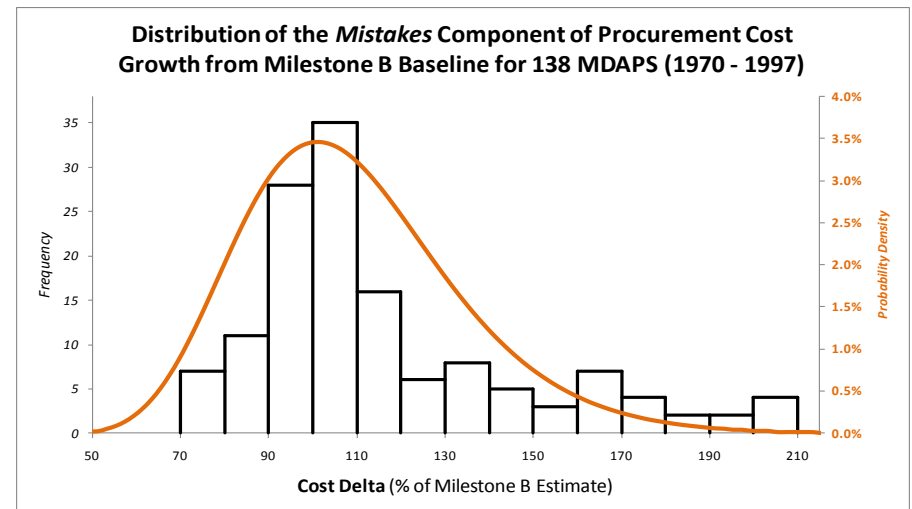
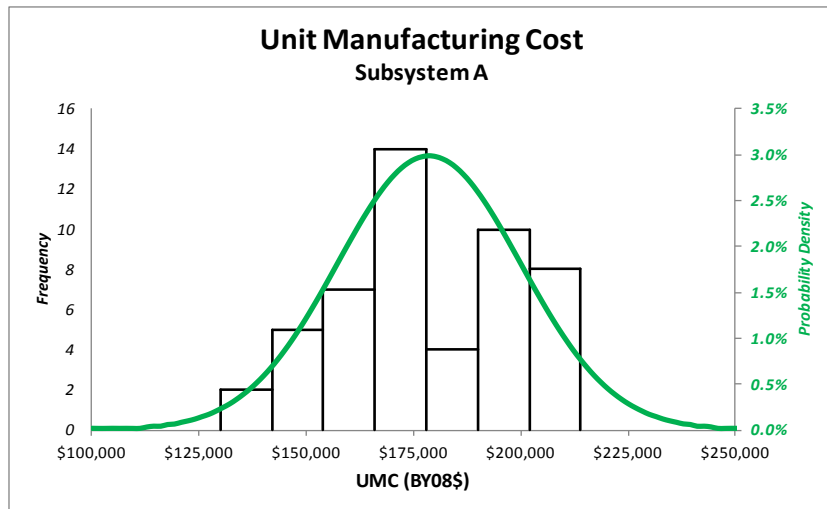




Accuracy Levels

How accurate could you possibly be?

- ▶ How much variation is there in our most reliable data sources?
- ▶ How well do we predict common variables (inflation, labor rates, fee, O/H, etc)?



“When predicting the price of a commodity as simple as a carton of eggs five years into the future, there is a standard error of 15%... Now imagine how much larger the standard error is for our sophisticated, state-of-the-art weapon systems that will take more than a decade to develop and procure.”

-COL Brian Shimel, USAF (*Nov-Dec 2008 Defense AT&L Magazine*)



The Unseen: Informing Decision Makers with Cost Risk Analysis



What Decision Makers Need to Know

- ▶ Confidence is “*perceived* probability”!
- ▶ What possibilities are extremely unlikely? Why do we view them to be so?
- ▶ What possibilities are most likely?
- ▶ ***The Unseen: What might happen that I’m not considering? Could I be on the path to disaster?***





TACOM

Life Cycle Management Command



Questions?