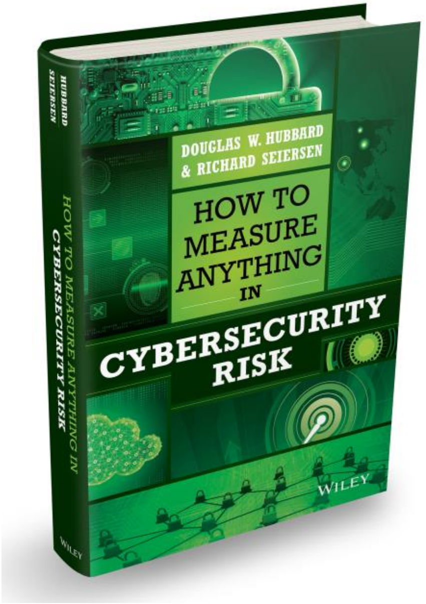
Hubbard Decision Research

How to Measure Anything in Cybersecurity Risk



Hubbard Decision Research
2 South 410 Canterbury Ct
Glen Ellyn, Illinois 60137
www.hubbardresearch.com

Introduction

My Co-Author and I

Richard Seiersen

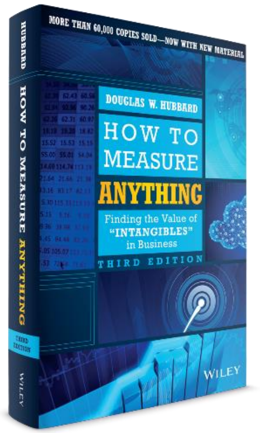
Currently the General Manager of Cybersecurity and Privacy at GE Health Care. Data driven executive with ~20 years experience spanning subject matters in Cyber Security, Quantitative Risk Management, Predictive Analytics, Big Data and Data Science, Enterprise Integrations and Governance Risk and Compliance (GRC). Led large enterprise teams, provided leadership in multinational organizations and tier one venture capital backed start-ups.

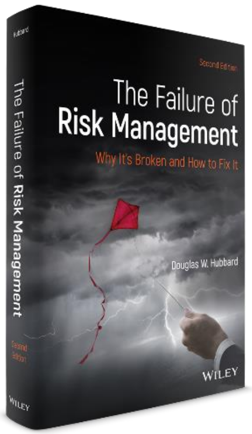
Douglas Hubbard

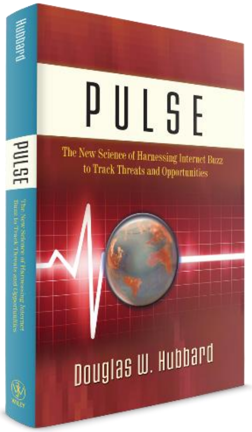
Mr. Hubbard is the inventor of the powerful Applied Information Economics (AIE) method. He is the author of the #1 bestseller in Amazon's math for business category for his book titled ***How to Measure Anything: Finding the Value of Intangibles in Business*** (Wiley, 2007; 3rd edition 2014). His other two books are titled ***The Failure of Risk Management: Why It's Broken and How to Fix It*** (Wiley, 2009) and ***Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities*** (Wiley, 2011).

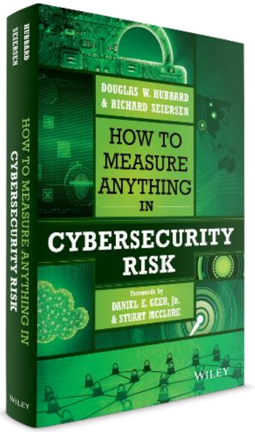
Introduction

My Books



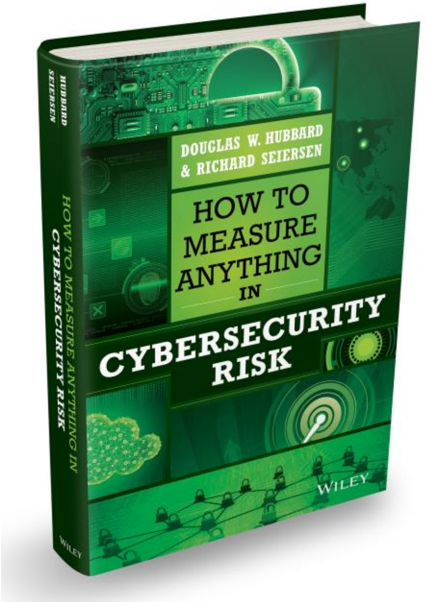






Introduction

How to Measure Anything in Cybersecurity Risk



" For thorough and practical guidance on using probability analysis for cybersecurity decision making, consult the book, How to Measure Anything in Cybersecurity "

Cite: CIS RAM Version 1.0 Center for Internet Security, Risk Assessment Method For Reasonable Implementation and Evaluation of CIS Controls

Introduction

Applied Information Economics

Applied Information Economics (AIE)

Information Technology

- Prioritizing IT portfolios
- Risk of software development
- Value of better information
- Value of better security
- Risk of obsolescence and optimal technology upgrades
- Value of network infrastructure
- Performance metrics for the business value of applications

Business Investments

- Prioritizing R&D in aerospace, biotech, pharma, medical devices and more
- Publishing
- Real estate
- Movie/film project selection

Engineering

- Power and road infrastructure upgrades
- Mining Risks

Government & Non-Profit

- Environmental policy
- Sustainable agriculture
- Procurement methods
- Grants management
- Public schools

Military

- Forecasting battlefield fuel consumption
- Effectiveness of combat training to reduce roadside bomb/IED casualties
- Methods for testing equipment

Introduction

The Biggest Cybersecurity Risk

Question: What is your single biggest risk in cybersecurity?

Answer: How you measure cybersecurity risk.

(This also applies to risk in general.)

Introduction

Topics for Today

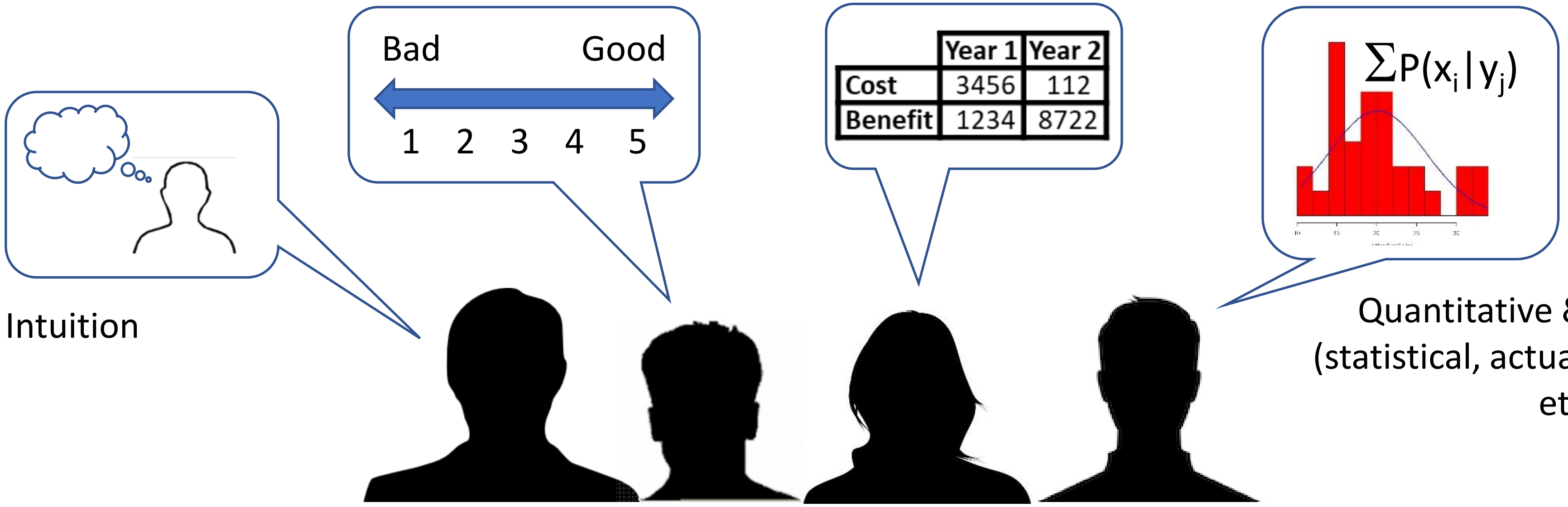
- What is wrong with current methods
- Why there are no immeasurables
- Improving the performance of experts
- Improving models with empirical data

Introduction

Types of Measurement Methods

Qualitative
(soft scores or “high/medium/low”)

Accounting-style
Cost estimate analysis
(point estimates, deterministic)

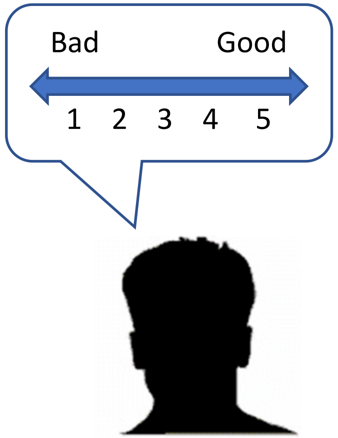


Expert Intuition

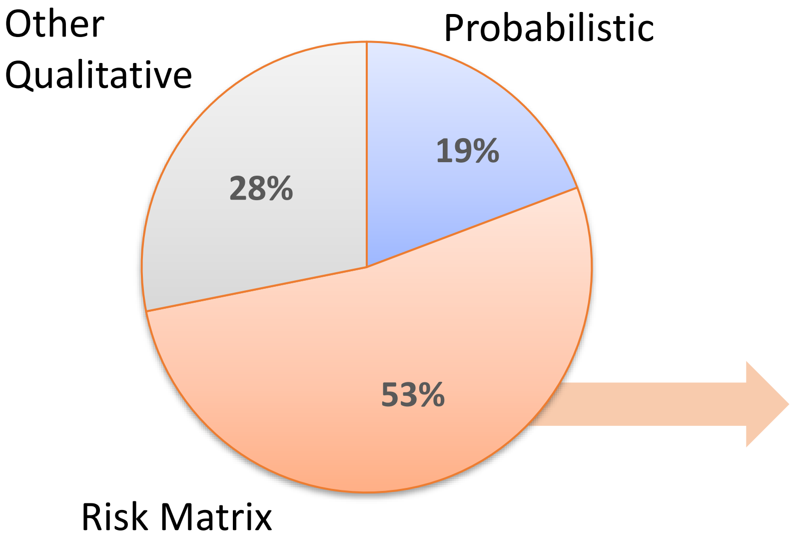
Quantitative & Probabilistic
(statistical, actuarial, simulations, etc.)

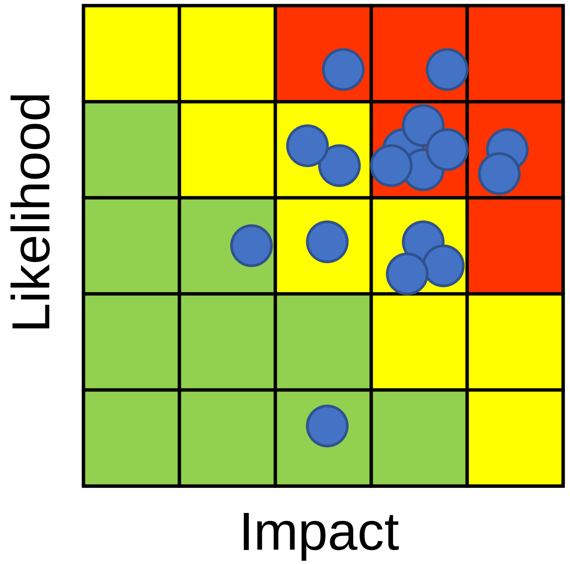
Do “Scores” and “Scales” Work?

The Current Most Popular Method



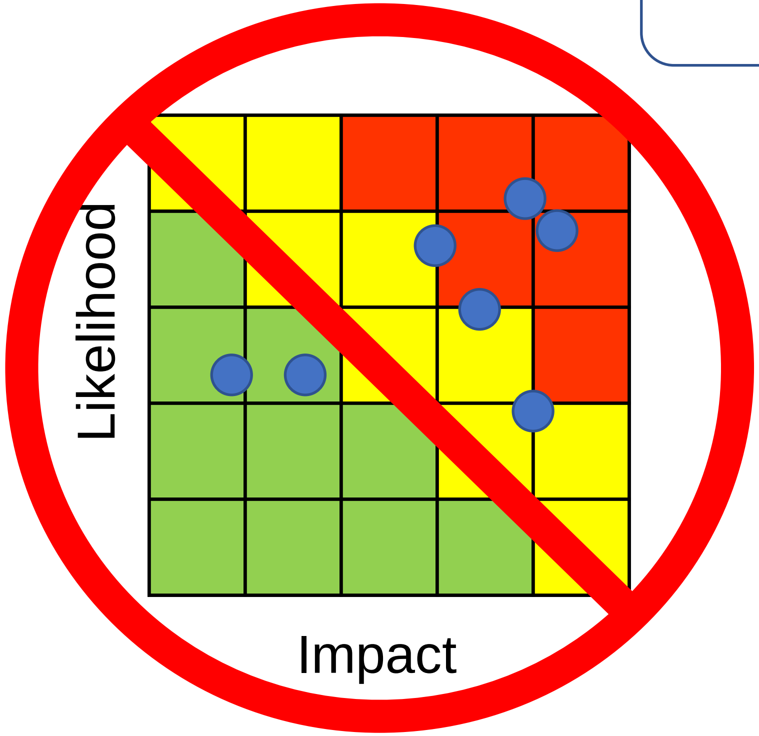
Share of Methods Used in Cybersecurity Risk Assessment





Do “Scores” and “Scales” Work?

The Ubiquitous Risk Matrix



“[Risk Matrices] can be worse than useless”

Risk Analysis 28, no. 2 (2008).

What’s Wrong with Risk Matrices?

L. A. Cox, Jr.

“Risk Matrices should not be used for decisions of any consequence”

Society of Petroleum Engineers Economics &Management 6, no. 2 (April

The Risk of Using Risk Matrices

P. Thomas, R. Bratvold, and J. E. Bickel

Abstract

The risk matrix (RM) is a widely espoused approach to assess and analyze risks in the oil & gas (O&G) industry. RMs have been implemented throughout that industry and are extensively used in risk-management contexts. This is evidenced by numerous SPE papers documenting RMs as the primary risk management tool. Yet, despite this extensive use, the key question remains to be addressed: Does the use of RMs guide us to make optimal (or even better) risk-management decisions?

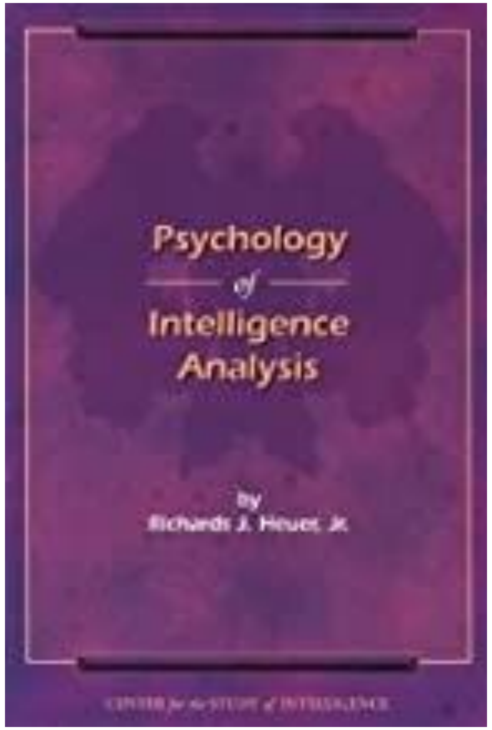
Do “Scores” and “Scales” Work?

How do we know what works?

“Intelligence analysts should be self-conscious about their reasoning processes. They should think about how they make judgments and reach conclusions, not just about the judgments and conclusions themselves.”

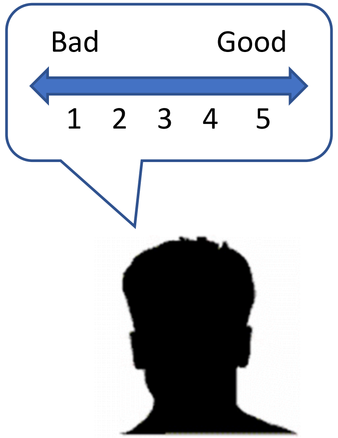
Dick Heuer, *The Psychology of Intelligence Analysis*

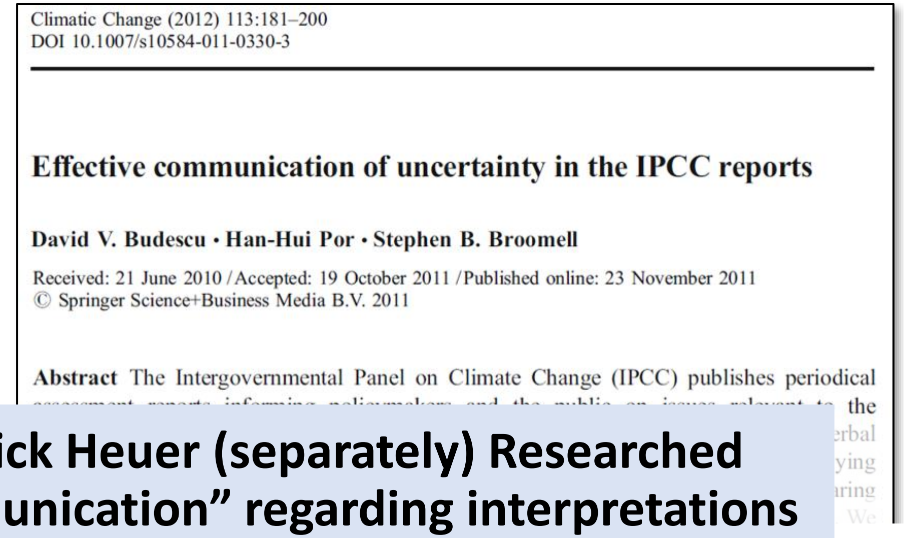
Meta-Decision Criteria: Is there real evidence, scientifically measured, that shows that one method is better than another?



Do “Scores” and “Scales” Work?

Unintended consequences of simple scoring methods





Climatic Change (2012) 113:181–200
DOI 10.1007/s10584-011-0330-3

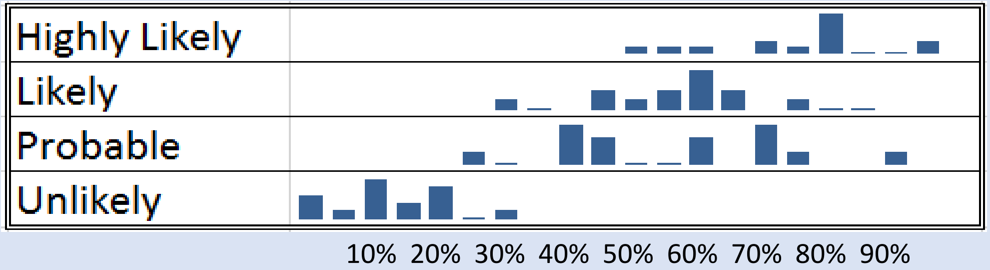
Effective communication of uncertainty in the IPCC reports

David V. Budescu • Han-Hui Por • Stephen B. Broomell

Received: 21 June 2010 / Accepted: 19 October 2011 / Published online: 23 November 2011

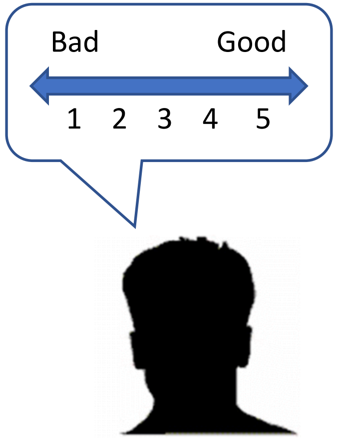
Abstract The Intergovernmental Panel on Climate Change (IPCC) publishes periodical

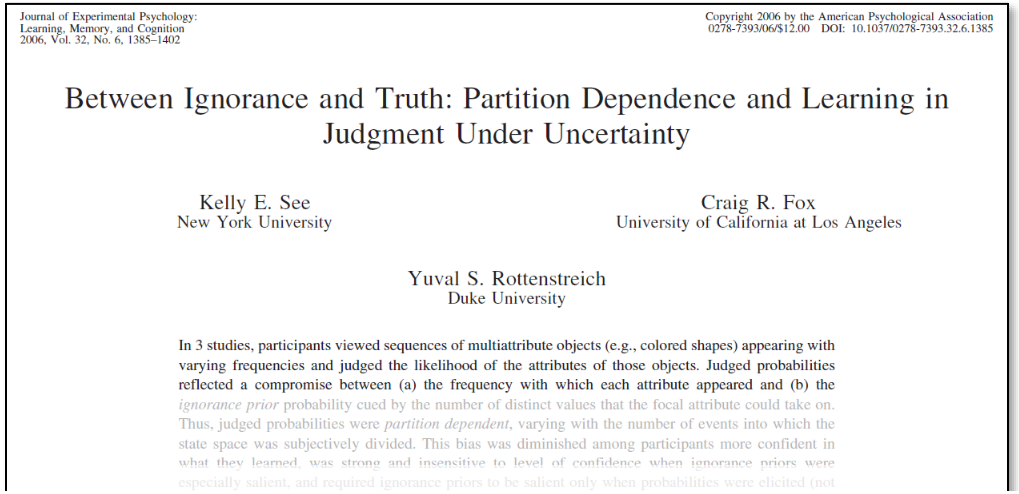
David Budescu and Dick Heuer (separately) Researched the “illusion of communication” regarding interpretations of verbal labels for probabilities



Do “Scores” and “Scales” Work?

Unintended consequences of simple scoring methods




Journal of Experimental Psychology: Learning, Memory, and Cognition
2006, Vol. 32, No. 6, 1385–1402

Copyright 2006 by the American Psychological Association
0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.6.1385

Between Ignorance and Truth: Partition Dependence and Learning in Judgment Under Uncertainty

Kelly E. See
New York University

Craig R. Fox
University of California at Los Angeles

Yuval S. Rottenstreich
Duke University

In 3 studies, participants viewed sequences of multiattribute objects (e.g., colored shapes) appearing with varying frequencies and judged the likelihood of the attributes of those objects. Judged probabilities reflected a compromise between (a) the frequency with which each attribute appeared and (b) the *ignorance prior* probability cued by the number of distinct values that the focal attribute could take on. Thus, judged probabilities were *partition dependent*, varying with the number of events into which the state space was subjectively divided. This bias was diminished among participants more confident in what they learned, was strong and insensitive to level of confidence when ignorance priors were especially salient, and required ignorance priors to be salient only when probabilities were elicited (not

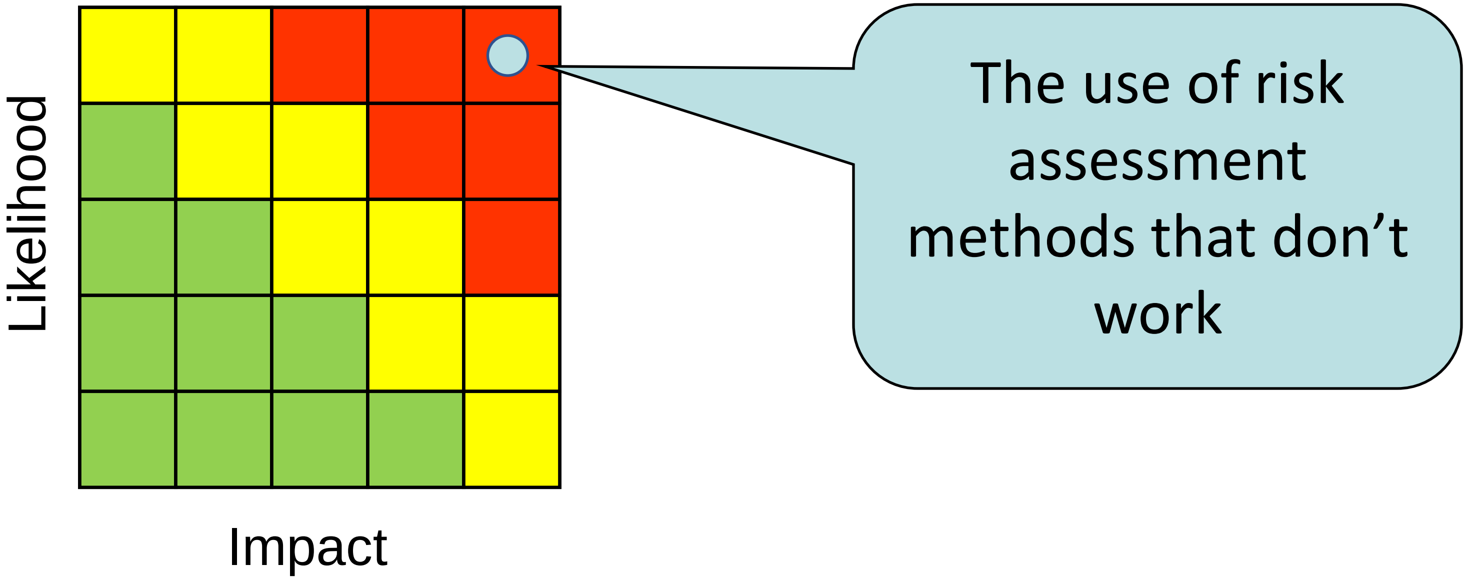
Craig R. Fox showed how arbitrary features of how scales are partitioned effects responses.

Example:

If “1” on a 5-point impact scale means “less than $1 million loss”, the share of that response is affected by the partition of *other* choices.

Do “Scores” and “Scales” Work?

The Only Risk Matrix You Need



The Analysis Placebo

Confidence in decision making methods is detached from performance

Organizational Behavior and Human Decision Processes 107, no. 2 (2008): 97– 105

Journal of Behavioral Decision Making 3, no. 3 (July/ September 1990): 153– 174

Law and Human Behavior 23 (1999): 499– 516.

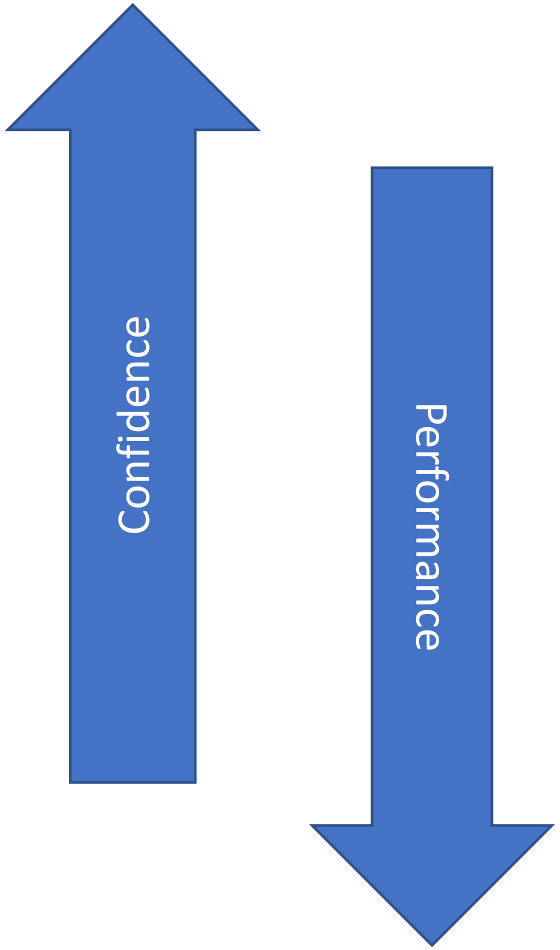
Organizational Behavior and Human Decision Processes 61, no. 3 (1995): 305– 326.

Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making

Heath and Gonzalez

Abstract

We present three studies of *interactive decision making,* where decision makers interact with others before making a final decision alone. Because the theories of lay observers and social psychologists emphasize the role of information collection in interaction, we developed a series of tests of information collection. Two studies



The Meta Decision

How to Build a Method That Works

- Start with components that work.
- Don’t rely on anecdotes, testimonials or claims of “best practices” as evidence of working.
- If you can’t answer “What is the probability of losing more than X in the next 12 months due to event Y?” then you aren’t doing risk analysis.

A Cybersecurity Survey

2015 Survey: Interesting Connection

Those who said they could "compute the probability of various levels of losses" had about *half the rate of data breaches* as those who could not.

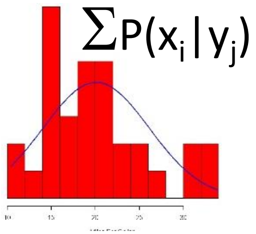
Does your organization compute the probability of various levels of losses?	Average Annual Data Breach Rate
Yes	4.5%
No	9%

173 responses total

A single survey might still be inconclusive – but it is consistent with other research about the improvement from using quantitative methods.

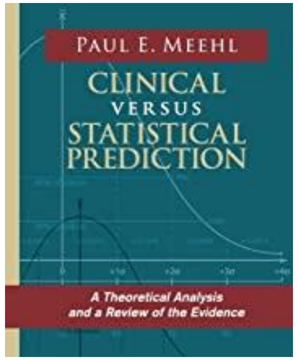
Experts vs. Algorithms

What the research says about statistical methods vs. Subject Matter Experts

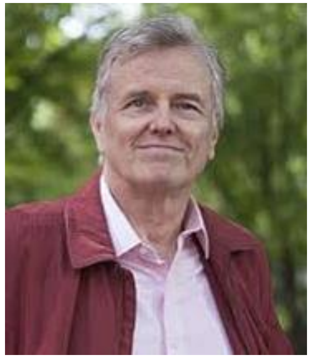


Paul Meehl assessed 150 studies comparing experts to statistical models in many fields (sports, prognosis of liver disease, etc.).

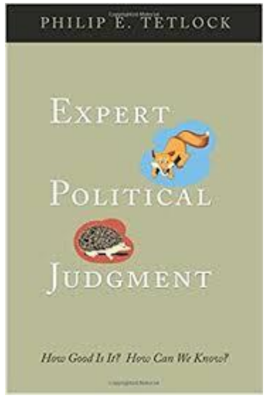
"There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one."



Philip Tetlock tracked a total of over 82,000 forecasts from 284 experts in a 20-year study covering politics, economics, war, technology trends and more.



"It is impossible to find any domain in which humans clearly outperformed crude extrapolation algorithms, less still sophisticated statistical ones."



What Measuring Risk Looks Like

Is Risk Analysis Actually Supporting Decisions?

- If risks and mitigation strategies were quantified in a meaningful way, decisions could be supported.
- In order to compute an ROI on mitigation decisions, we need to quantify likelihood, monetary impact, cost, and effectiveness.

	Expected Loss/Yr	Cost of Control	Control Effectiveness	Return on Control	Action
DB Access	$24.7M	$800K	95%	2,832%	Mitigate
Physical Access	$2.5M	$300K	99%	727%	Mitigate
Data in Transit	$2.3M	$600K	95%	267%	Mitigate
Network Access Control	$2.3M	$400K	30%	74%	Mitigate
File Access	$969K	$600K	90%	45%	Monitor
Web Vulnerabilities	$409K	$800K	95%	-51%	Track
System Configuration	$113K	$500K	100%	-77%	Track

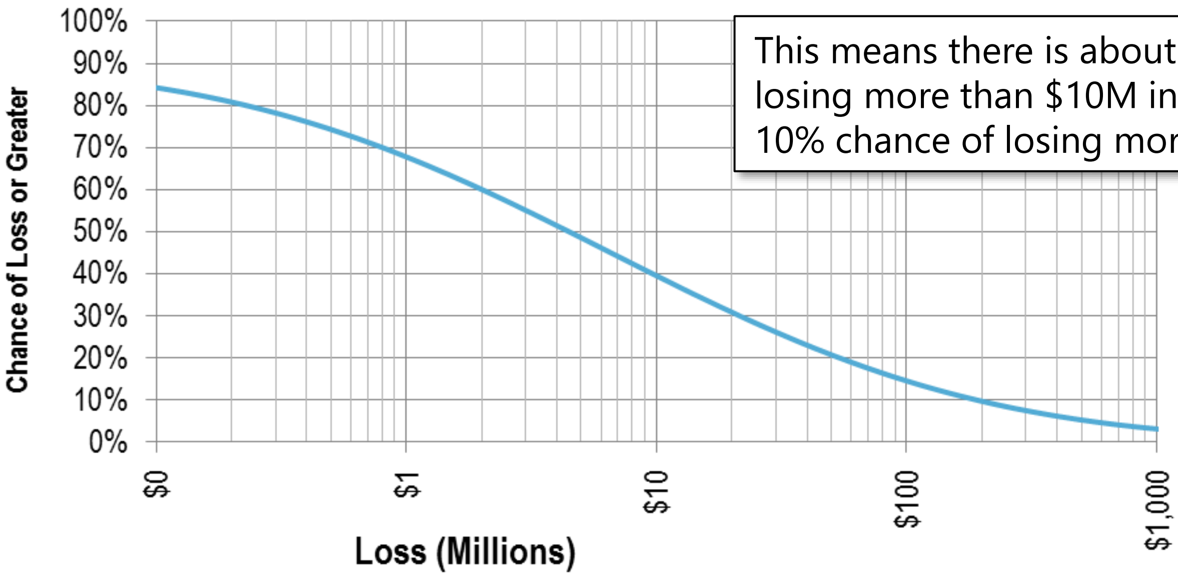
What Measuring Risk Looks Like

The Loss Exceedance Curve

What if we could measure risk more like an actuary? For example, "The probability of losing more than $10 million due to security incidents in 2016 is 16%."

What if we could prioritize security investments based on a "Return on Mitigation"?

	Expected Loss/Yr	Cost of Control	Control Effectiveness	Return on Control	Action
DB Access	$24.7M	$800K	95%	2,832%	Mitigate
Physical Access	$2.5M	$300K	99%	727%	Mitigate
Data in Transit	$2.3M	$600K	95%	267%	Mitigate
Network Access Control	$2.3M	$400K	30%	74%	Mitigate
File Access	$969K	$600K	90%	45%	Monitor
Web Vulnerabilities	$409K	$800K	95%	-51%	Track
System Configuration	$113K	$500K	100%	-77%	Track

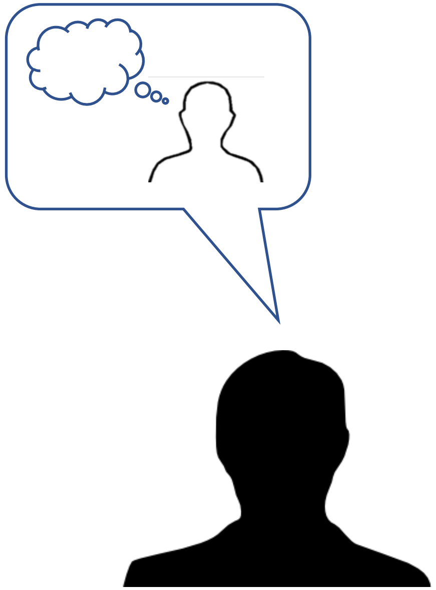


This means there is about a 40% chance of losing more than $10M in a year and about a 10% chance of losing more than $200M.

The Method of Measurement

Why Does Our Risk Tolerance Change?

Decision makers are also inconsistent regarding their own aversion to risk.



Neuron Vol. 47, (2005): 763–770

The Neural Basis of Financial Risk Taking

Camelia M. Kuhnen and Brian Knutson

Journal of Personality and Social Psychology
2001, Vol. 81, No. 1, 146–159

0022-3514/01/$5.00 DOI: 10.1037//0022-3514.81.1.146

Fear, Anger, and Risk

Jennifer S. Lerner
Carnegie Mellon University

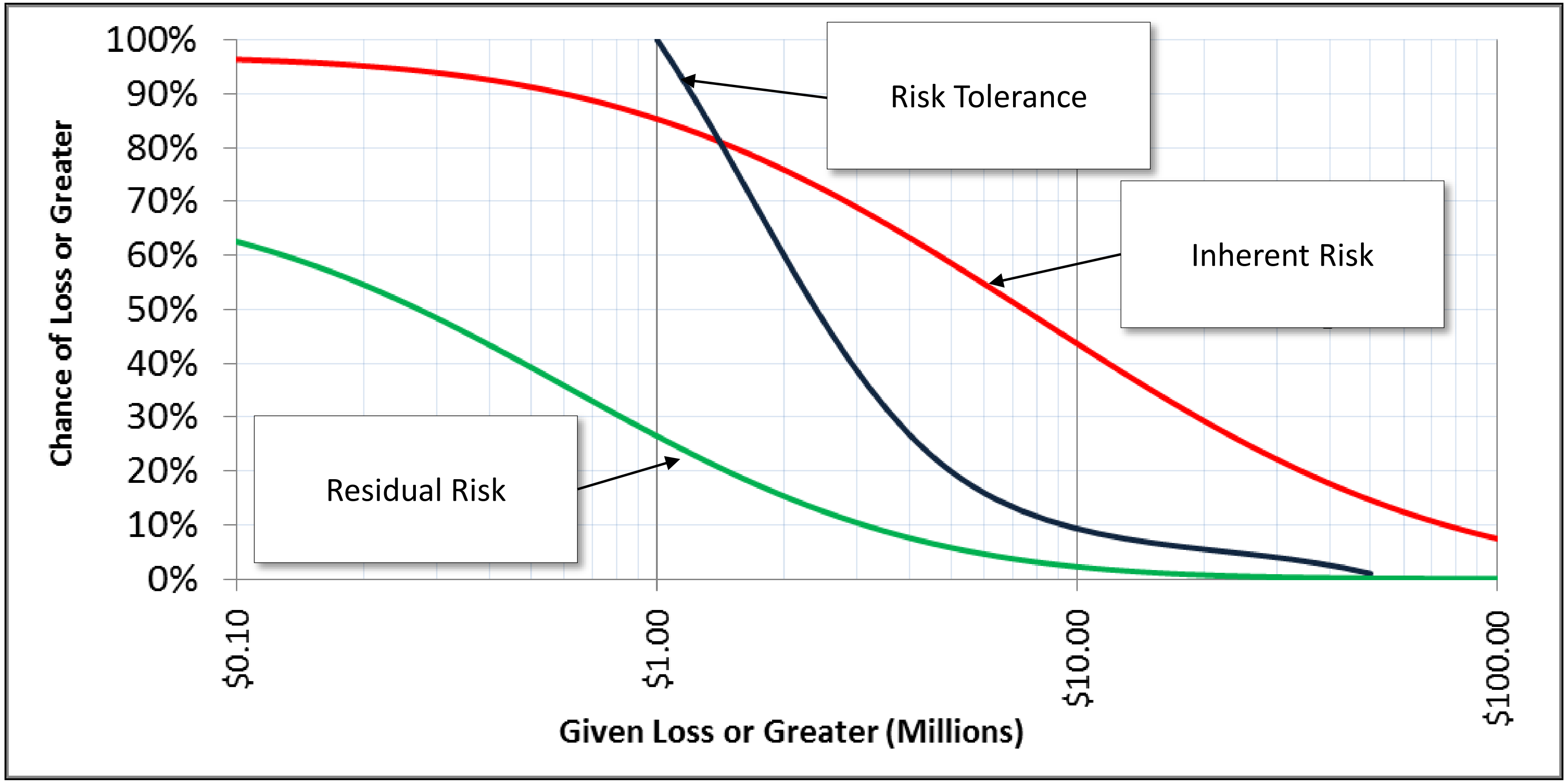
Dacher Keltner
University of California, Berkeley

Factor	Risk Aversion
Being around smiling people	↓
Recalling an event causing fear	↑
Recalling an event causing anger	↓
A recent win in an unrelated decision	↓
A recent loss in an unrelated decision	↑

Measuring Risk

Loss Exceedance Curves: Before and After

How do we show the risk exposure after applying available mitigations?



What Measuring Risk Looks Like

A Simple "One-For-One Substitution"

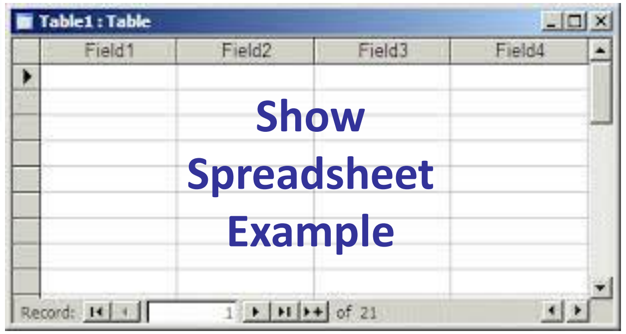
Each of these examples can be found on

www.howtomeasureanything.com/cybersecurity

Event	Event Probability (per Year)	Impact (90% Confidence Interval)		Random Result (zero when the event did not occur)
		Lower Bound	Upper Bound	
AA	.1	$50,000	$500,000	0
AB	.05	$100,000	$10,000,000	$8,456,193
AC	.01	$200,000	$25,000,000	0
AD	.03	$100,000	$15,000,000	0
AE	.05	$250,000	$30,000,000	0
AF	.1	$200,000	$2,000,000	0
AG	.07	$1,000,000	$10,000,000	$2,110,284
AH	.02	$100,000	$15,000,000	0
ZM	.05	$250,000	$30,000,000	0
ZN	.01	$1,500,000	$40,000,000	0
			Total:	$23,345,193

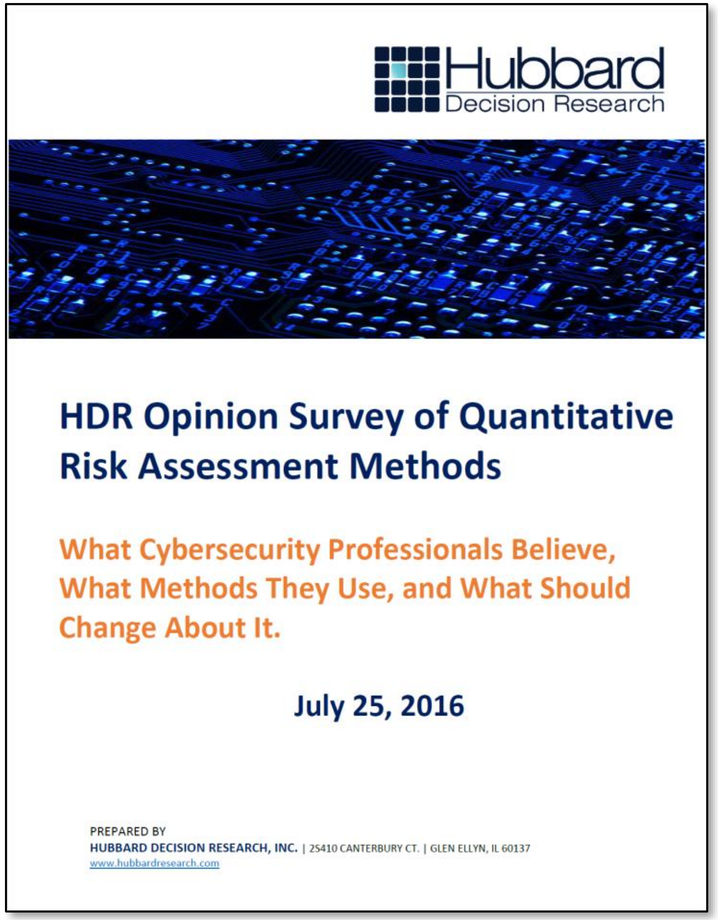
Each "Dot" on a risk matrix can be better represented as a row on a table like this

The output can then be represented as a Loss Exceedance Curve.



Obstacles to Better Decisions

Acceptance of quantitative methods vs. statistical literacy: survey results



Hubbard
Decision Research

HDR Opinion Survey of Quantitative Risk Assessment Methods

What Cybersecurity Professionals Believe, What Methods They Use, and What Should Change About It.

July 25, 2016

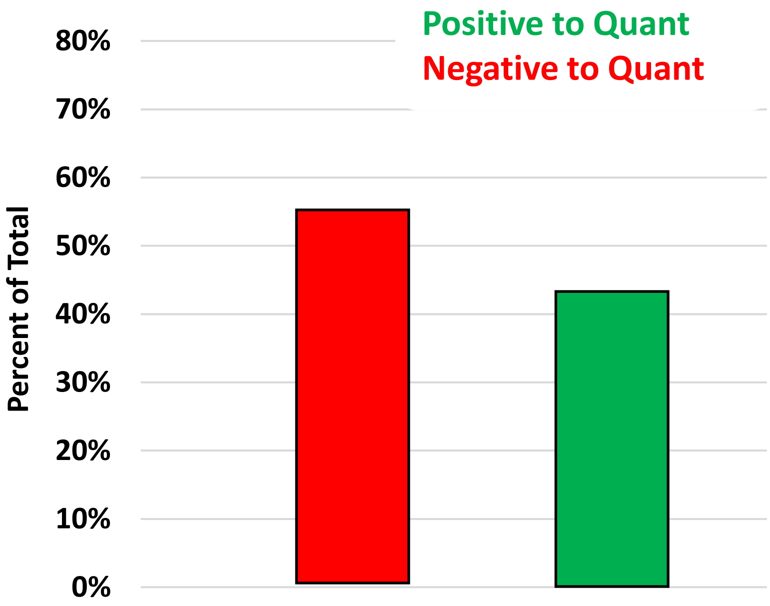
PREPARED BY
HUBBARD DECISION RESEARCH, INC. | 25410 CANTERBURY CT. | GLEN ELLYN, IL 60137
www.hubbardresearch.com

- 173 cybersecurity were surveyed regarding opinions about quantitative risk analysis methods in their fields
- There was a bit more resistance to quantitative methods than acceptance.
- They also took a quiz on basic statistical literacy
- When we looked only at those responses that scored above the median on statistical literacy, there was a lot more acceptance.
- When we look at those that did not score above the median, resistance was much higher.
- Those who answered “I don’t know” on stats literacy questions were not the most resistant to quantitative methods – it was those who thought they did know and were wrong.

So Why Don’t We Use More Quantitative Methods?

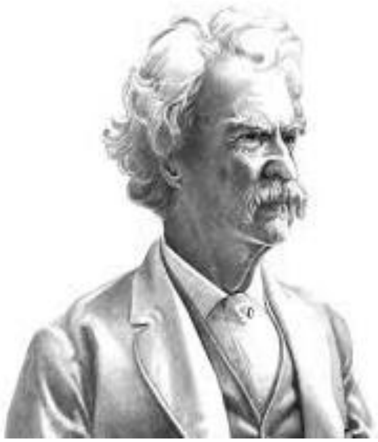
The Main Obstacle to Quantitative Methods

Another finding in the same survey: Strong opinions against “quant” are associated with poor stats understanding.



“It’s not what you don’t know that will hurt you, it’s what you know that ain’t so.”

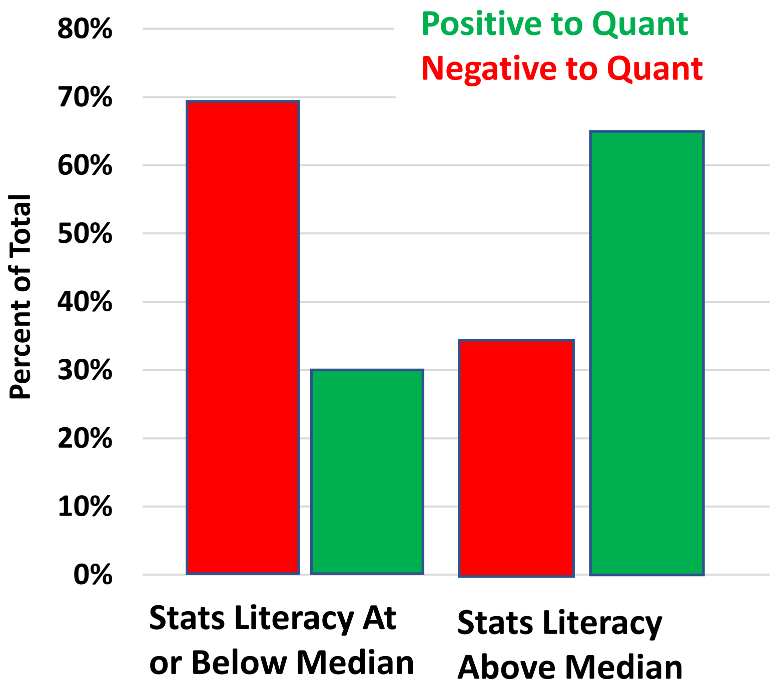
Mark Twain



So Why Don't We Use More Quantitative Methods?

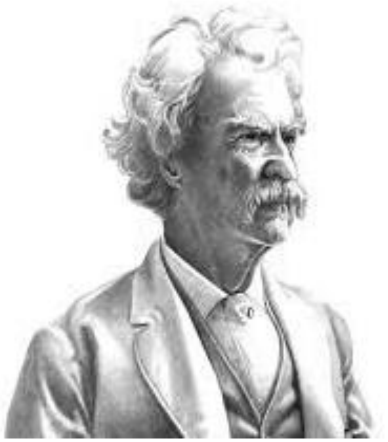
The Main Obstacle to Quantitative Methods

Another finding in the same survey: Strong opinions against "quant" are associated with poor stats understanding.



"It's not what you don't know that will hurt you, it's what you know that ain't so."

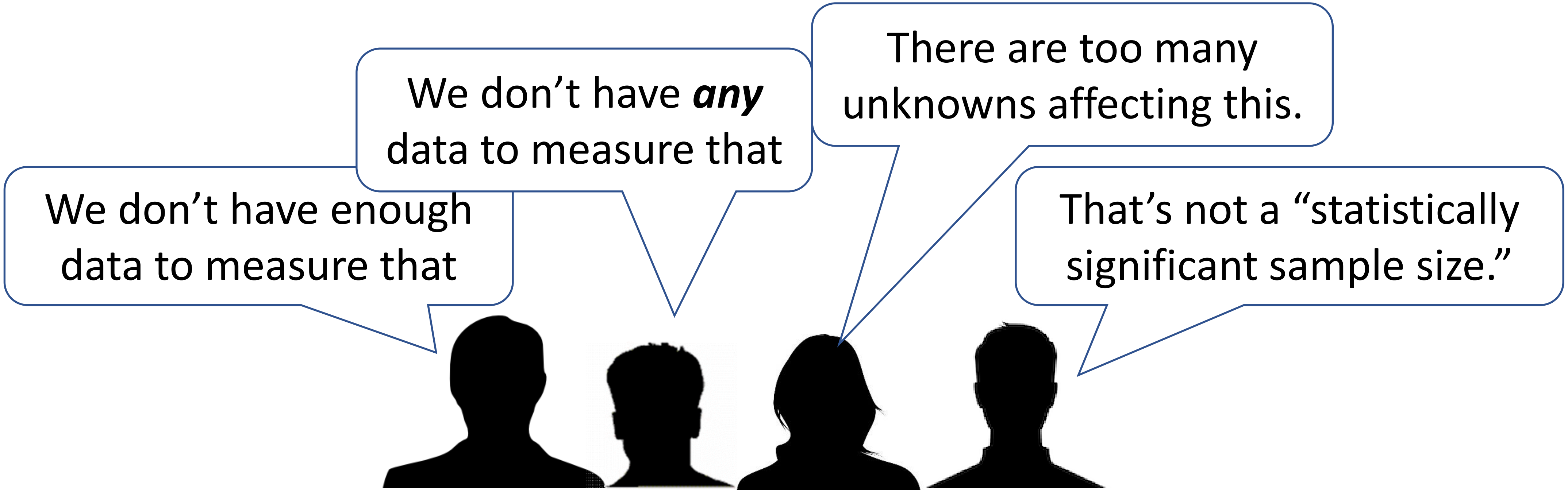
Mark Twain



So Why Don’t We Use More Quantitative Methods?

Commonly stated reasons for not using quantitative methods

Have you heard (or said) any of these?



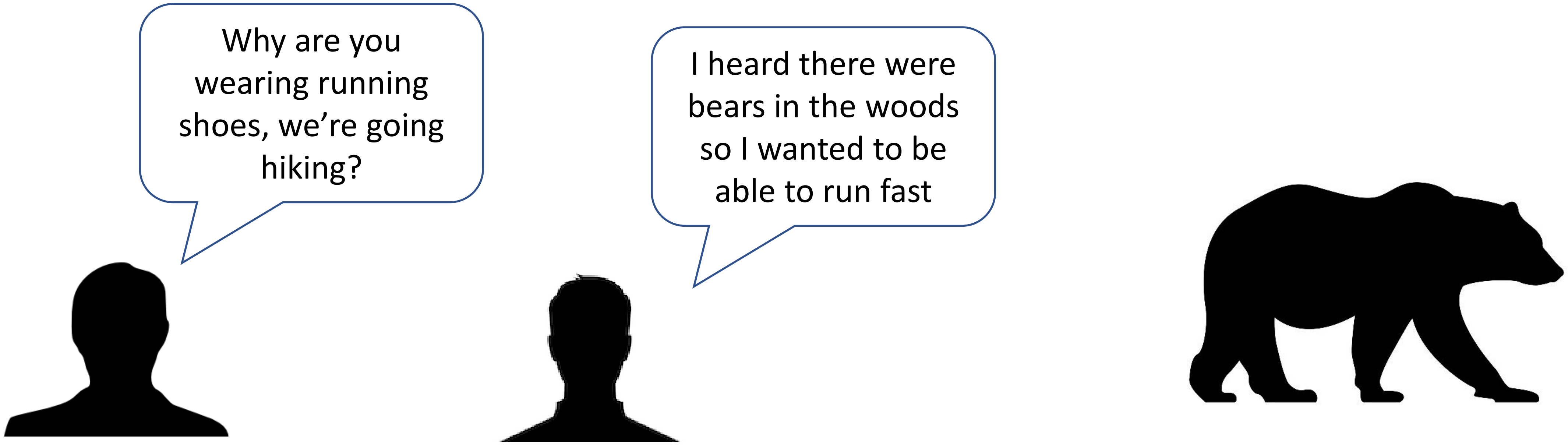
So Why Don’t We Use More Quantitative Methods?

Commonly stated reasons for not using quantitative methods

The implied (and unjustified) conclusion from each of these is….

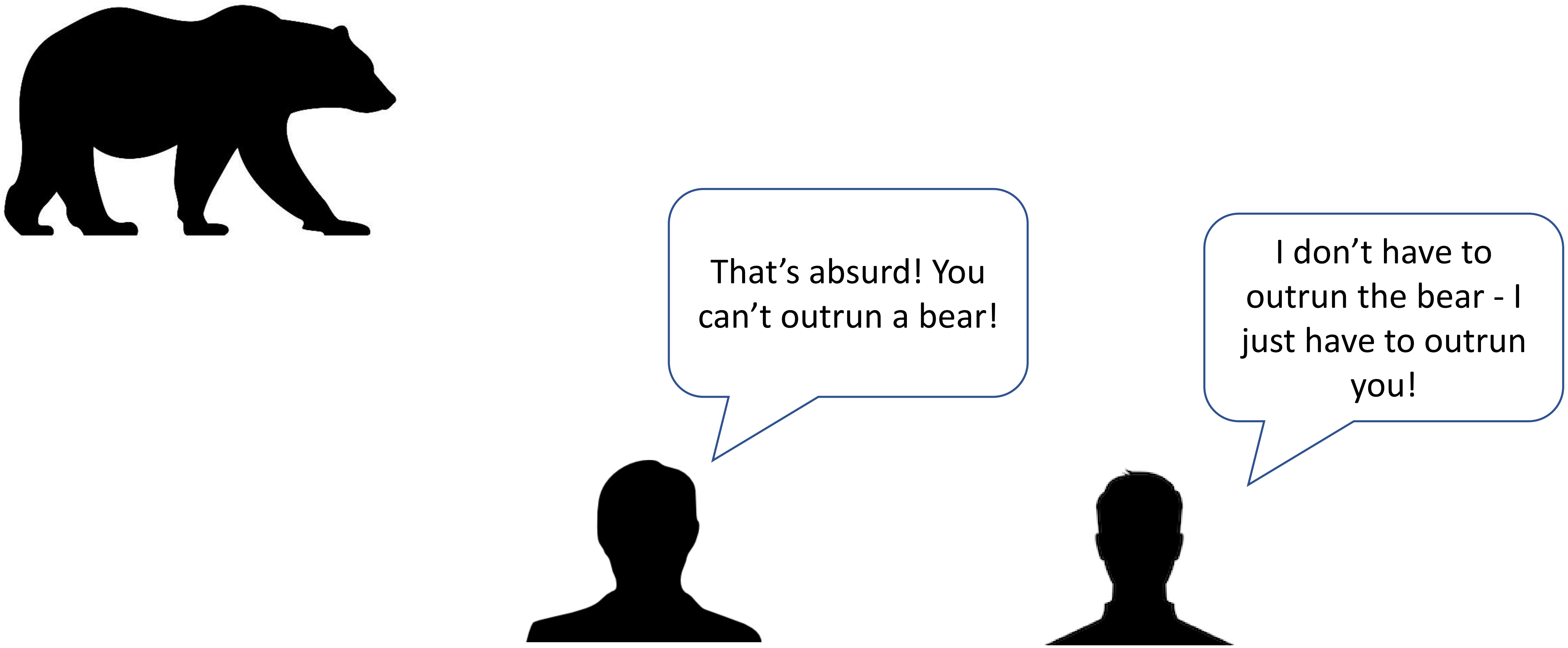
Obstacles to Better Decisions

The Double Standard



Obstacles to Better Decisions

The Double Standard

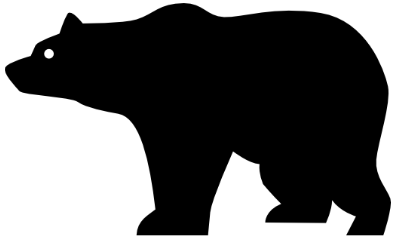


Irrational Bias Against Algorithms

A Double Standard

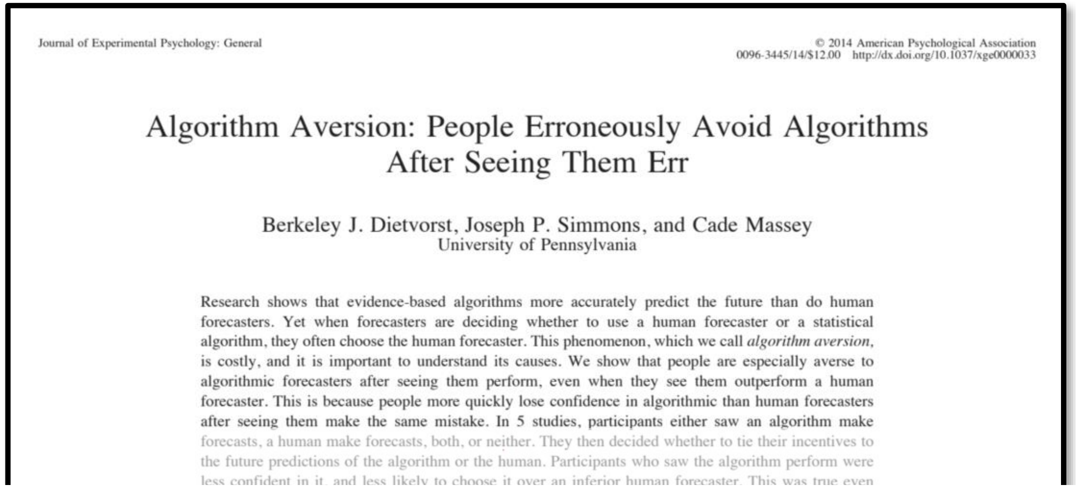
Don't commit the classic
"Beat the Bear" fallacy.

Exsupero Ursus



A common form of the ***Exsupero Ursus*** fallacy:
"The quantitative model must have
1) All the variables
2) All the data
3) All the right distributions and correlations
4) All the above

If not, default to a measurably worse method.



Journal of Experimental Psychology: General

© 2014 American Psychological Association
0096-3445/14/$12.00 http://dx.doi.org/10.1037/xge0000033

Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err

Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey
University of Pennsylvania

Research shows that evidence-based algorithms more accurately predict the future than do human forecasters. Yet when forecasters are deciding whether to use a human forecaster or a statistical algorithm, they often choose the human forecaster. This phenomenon, which we call *algorithm aversion*, is costly, and it is important to understand its causes. We show that people are especially averse to algorithmic forecasters after seeing them perform, even when they see them outperform a human forecaster. This is because people more quickly lose confidence in algorithmic than human forecasters after seeing them make the same mistake. In 5 studies, participants either saw an algorithm make forecasts, a human make forecasts, both, or neither. They then decided whether to tie their incentives to the future predictions of the algorithm or the human. Participants who saw the algorithm perform were less confident in it, and less likely to choose it over an inferior human forecaster. This was true even

The Three Misconceptions Behind Any Perceived "Immeasurable"

The Illusions of Immeasurability

CONCEPT of Measurement	The definition of measurement itself is widely misunderstood.
OBJECT of Measurement	The thing being measured is not well defined.
METHOD of Measurement	Many procedures of empirical observation are misunderstood.

The Three Misconceptions Behind Any Perceived "Immeasurable"

The Concept of Measurement

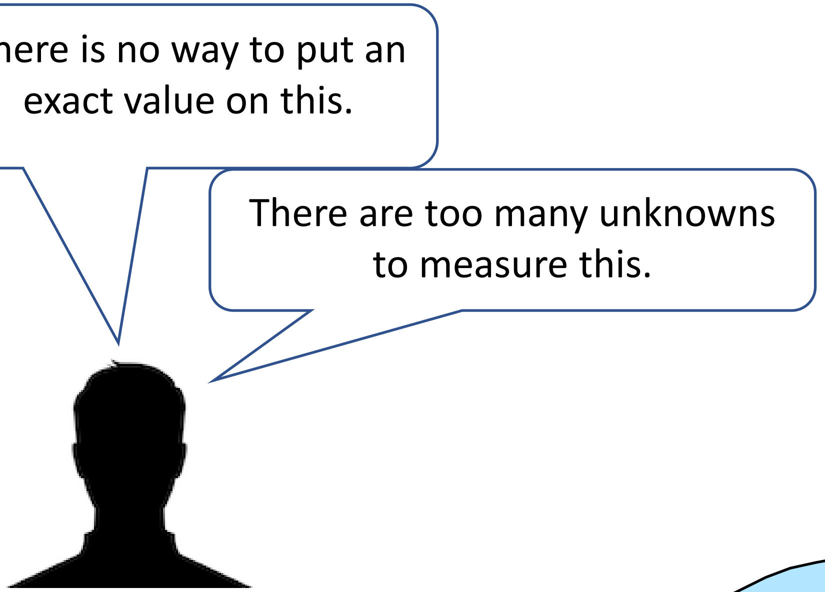
CONCEPT of Measurement	The definition of measurement itself is widely misunderstood.
OBJECT of Measurement	The thing being measured is not well defined.
METHOD of Measurement	Many procedures of empirical observation are misunderstood.

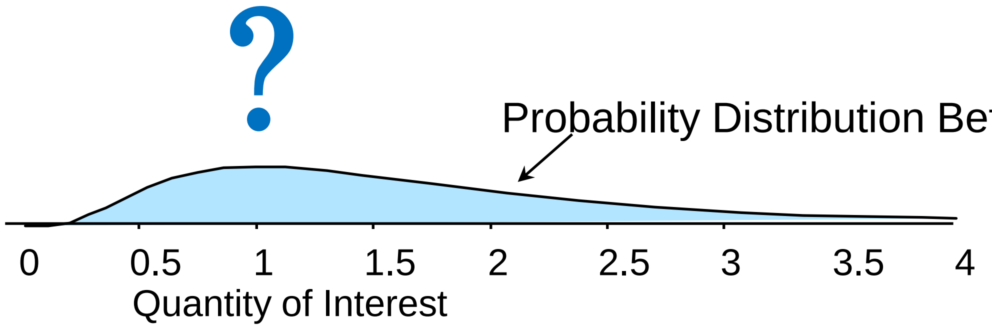
The Concept of Measurement

What Measurement Really Means

It's not a point value.

<u>Measurement:</u> a quantitatively expressed reduction in uncertainty based on observation.



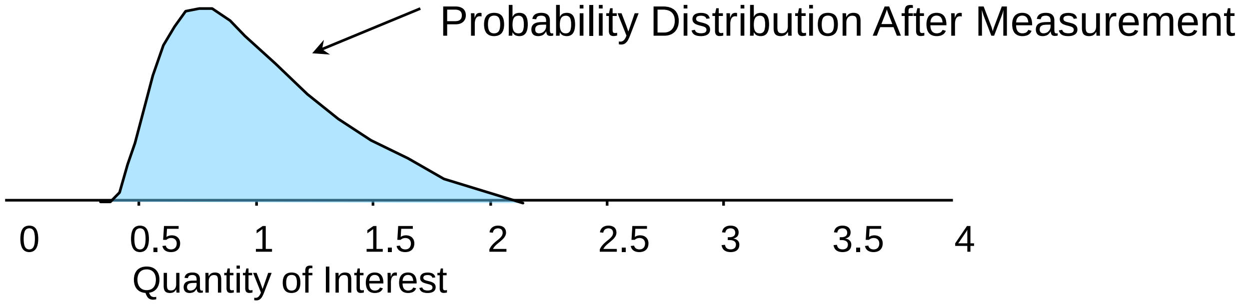


The Concept of Measurement

What Measurement Really Means

It’s not a point value.

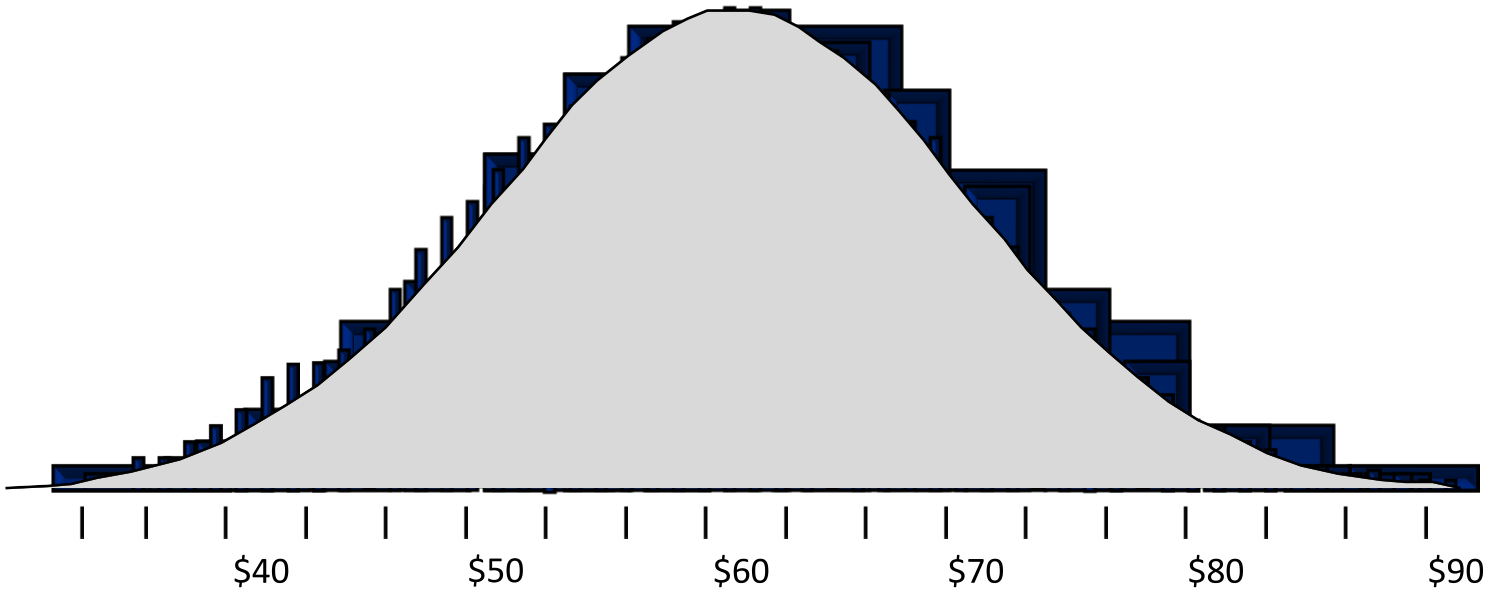
<u>Measurement:</u> a quantitatively expressed reduction in uncertainty based on observation.



The Concept of Measurement

Constructing a Distribution

- Uncertainty about “either/or” events are expressed as “discrete” probabilities (e.g. “35%).
- Uncertainty about continuous values can still be thought of as sets of discrete probabilities.



The Concept of Measurement

What the research says about Subject Matter Experts

"Overconfident professionals sincerely believe they have expertise, act as experts and look like experts. You will have to struggle to remind yourself that they may be in the grip of an illusion."

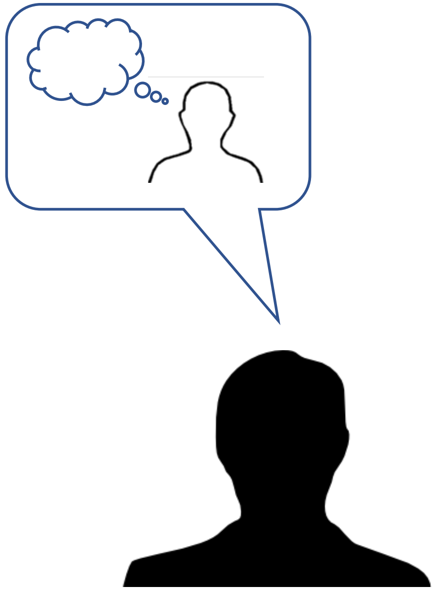
Daniel Kahneman, Psychologist, Economics Nobel

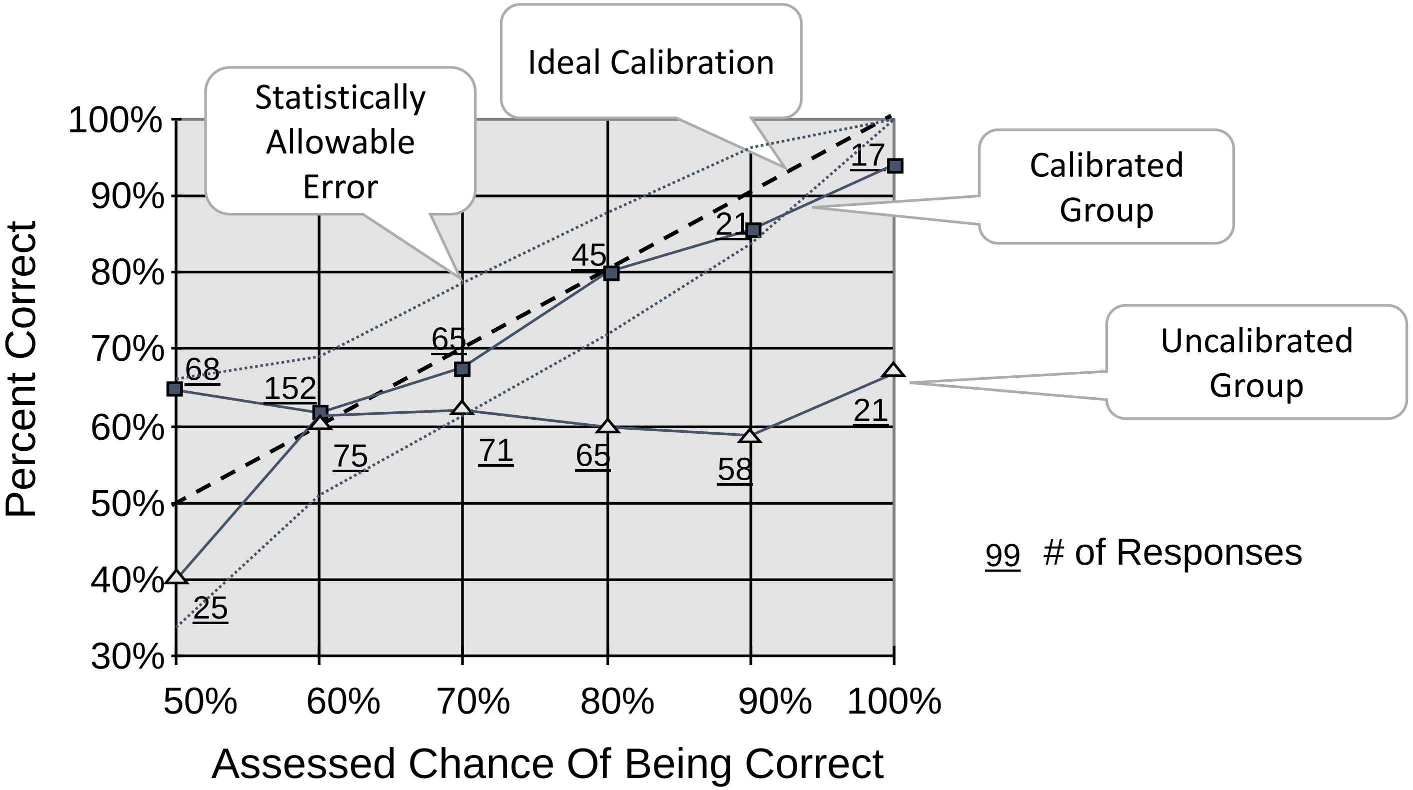
- Decades of studies show that most managers are statistically "overconfident" when assessing their own uncertainty.
- Studies also show that measuring *your own* uncertainty about a quantity is a general skill that <u>can be taught</u> with a ***measurable*** improvement.

The Concept of Measurement

Training Experts to Give Calibrated Probabilities

Training can “calibrate” people so that of all the times they say they are 90% confident, they will be right 90% of the time.



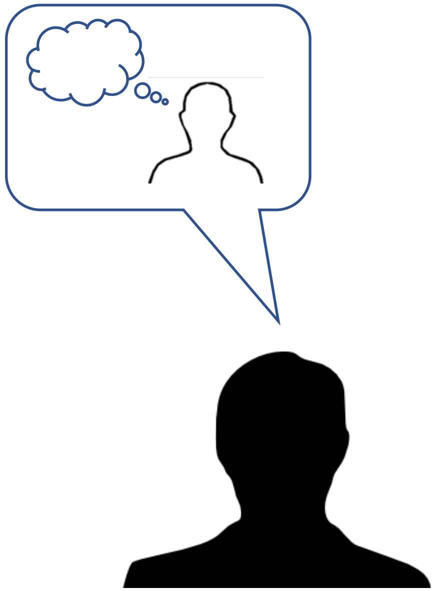


Source: Hubbard Decision Research, Giga Information Group

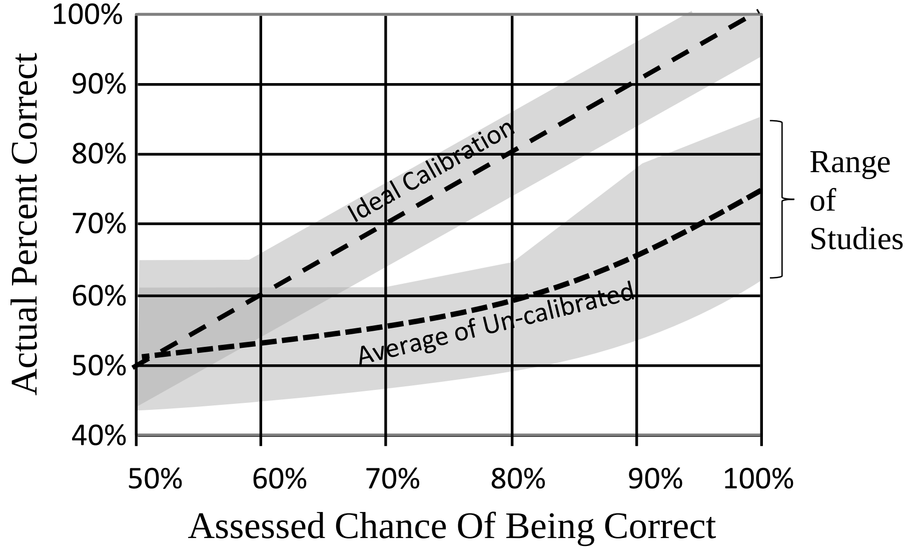
The Concept of Measurement

Overconfidence in Ranges

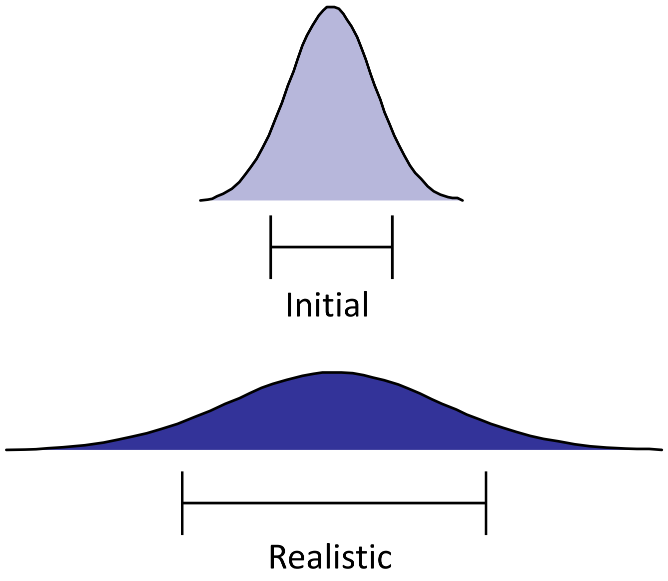
The same training methods apply to the assessment of uncertain ranges for quantities like the duration of a future outage, the records compromised in a future breach, etc.



Binary Events
(It happens or not, like a chance of data breach)



90% Confidence Interval
(For continuous values, like impact)



The Concept of Measurement

Calibration Exercise: Ranges

For the following questions, provide a range (an upper and lower bound) that you are 90% certain contains the correct answer:

Questions	Lower Bound	Upper Bound
Napoleon Bonaparte was born what year?		
What is the average weight of an adult male African elephant (tons)?		
The Coliseum in Rome held how many spectators?		
How many countries were in NATO in 2010?		
In what year did Newton publish the Laws of Gravitation?		

The Concept of Measurement

Calibration Exercise: True/False

For each statement below, answer whether you believe it is true or false and provide a percentage confidence that your answer is correct. Confidence is any value between 50% (“no idea”) to 100% (certainty).

Questions	True or False?	% Confidence
Brazil has a larger population than Spain.		
A hockey puck will fit in a golf hole.		
The Yangtze River is the longest river in Asia.		
Mars is always further away from Earth than Venus is from Earth.		
The movie *Titanic* still holds the record for box office receipts in the first six weeks.		

The Concept of Measurement

Calibration Answers

	Lower Bound
Napoleon Bonaparte was born what year?	1769
What is the average weight of an adult male African elephant (tons)?	3.5 tons
The Coliseum in Rome held how many spectators?	50,000
How many countries were in NATO in 2010?	28
In what year did Newton publish the Laws of Gravitation?	1687

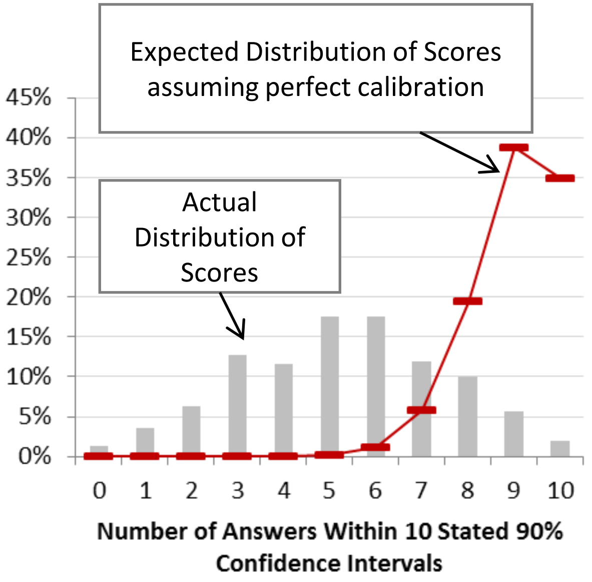
	True or False?
Brazil has a larger population than Spain.	True
A hockey puck will fit in a golf hole.	True
The Yangtze River is the longest river in Asia.	True
Mars is always further away from Earth than Venus is from Earth.	False
The movie Titanic still holds the record for box office receipts in the first six weeks.	False

The Concept of Measurement

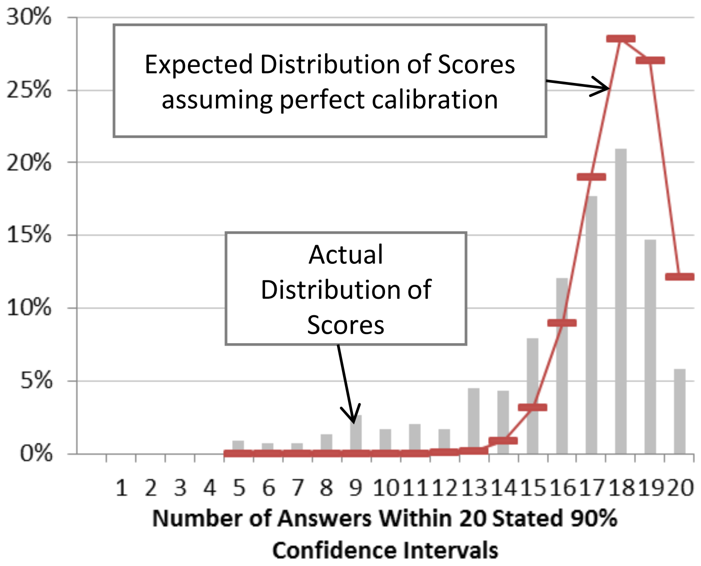
More Data on the Effects of Calibration Training

- With nearly 1,000 subjects who have taken the same calibration tests, and over 100,000 individual responses, HDR has more calibration data than all academic literature combined.
- A clear pattern emerges: Training has a major impact; 15% don’t quite reach calibration

Initial 10 Question 90% CI Test



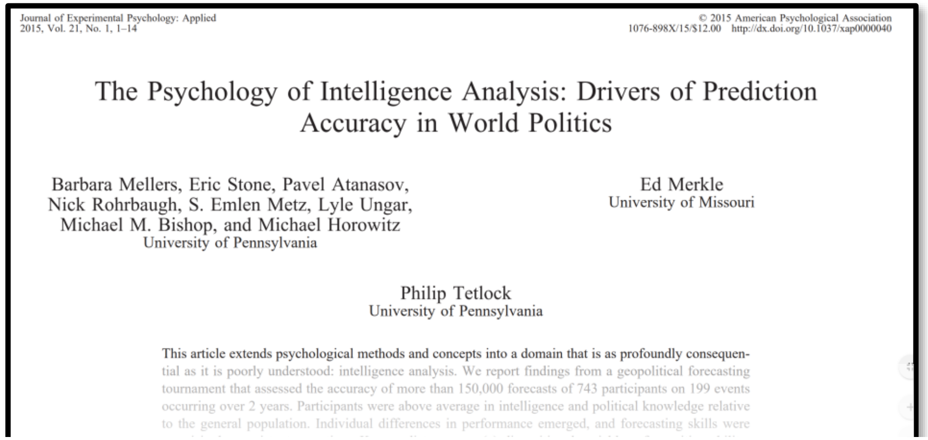
Final 20 Question 90% CI Test



The Concept of Measurement

Improving Expert Forecasts

- Tetlock also looked at what improved *forecasting.*
- He tracked 743 individuals who made at least 30 forecasts each over a 2-year period.
- He determined factors that made the biggest difference in the performance of forecasting.


Journal of Experimental Psychology: Applied
2015, Vol. 21, No. 1, 1–14

1076-898X/15/$12.00 http://dx.doi.org/10.1037/xap0000040

The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

Barbara Mellers, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were

Probabilistic Training

- Subjects were trained in basic inference methods, using reference classes, and avoiding common errors and biases.

Teams and Belief Updating

- Teams deliberated more and individuals were willing to update beliefs based on new information.

Selecting the Best

- Brains matter. Both topic expertise and overall IQ were the best predictors of performance.

The Three Misconceptions Behind Any Perceived "Immeasurable"

The Object of Measurement

CONCEPT of Measurement	The definition of measurement itself is widely misunderstood.
OBJECT of Measurement	The thing being measured is not well defined.
METHOD of Measurement	Many procedures of empirical observation are misunderstood.

The Object of Measurement

The Importance of Defining a Measurement

- If a thing seems like an immeasurable “intangible” it may just be ill-defined.
- Often, if we can define what we mean by a certain “intangible” we find ways to measure it.
- Examples: Brand image, Security, Safety, etc.

The Object of Measurement

Clarifying the Problem

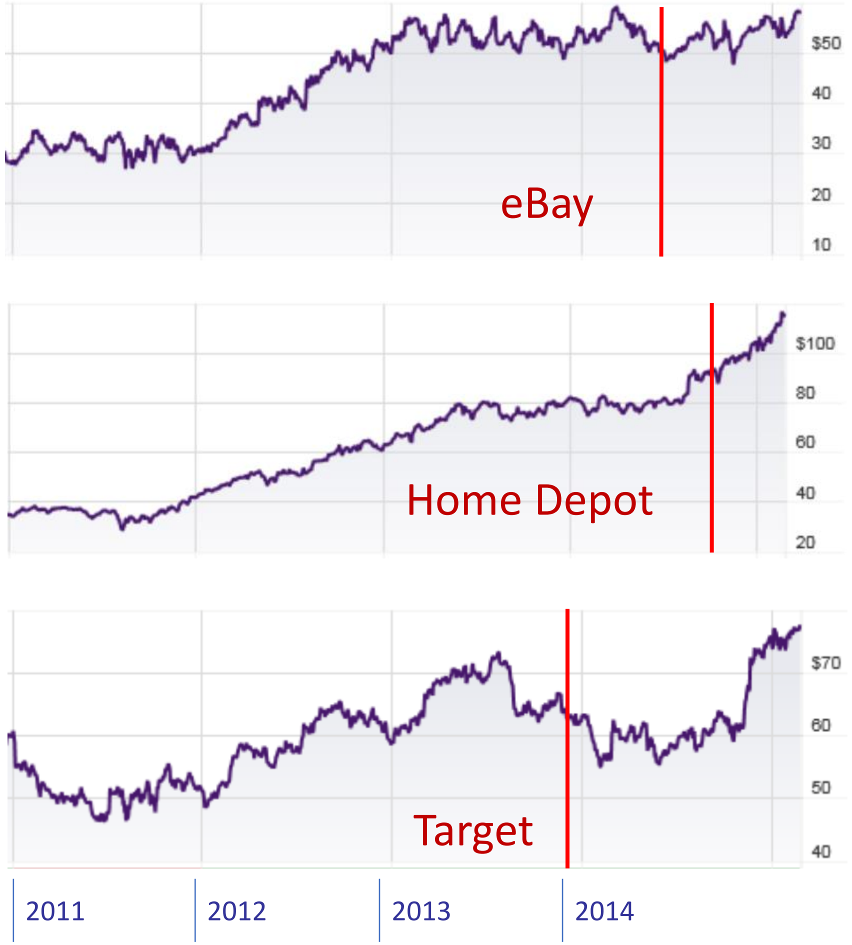
1. Why do you care? (What decision could depend on the outcome of this measurement?)
2. What do you see when you see more of it? (Describe it in terms of observable consequences, then units of measure.)
3. How much do you know about it now?
4. At what point will the value make a difference?
5. How much is additional information worth?

If you can answer the first three, you can usually compute the last two.

The Object of Measurement

Measurement Challenge: Reputation Damage

- One of the perceived most difficult measurements in cybersecurity is damage to reputation.
- Trick: *There is no such thing as a “secret” damage to reputation!*
- How about comparing stock prices after incidents? (That’s all public!)
- So what is the *REAL* damage?
    - Legal liabilities,
    - Customer outreach
    - “Penance” projects (security overkill)
- The upshot, damage to reputation actually has available information and easily observable measured costs incurred to *avoid* the bigger damages!



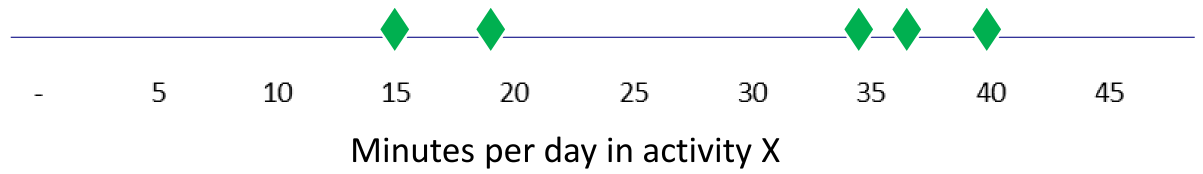
The Three Misconceptions Behind Any Perceived "Immeasurable"

The Method of Measurement

CONCEPT of Measurement	The definition of measurement itself is widely misunderstood.
OBJECT of Measurement	The thing being measured is not well defined.
METHOD of Measurement	Many procedures of empirical observation are misunderstood.

A Sample of 5

- Suppose you are extremely uncertain about how much time per day is spent in some activity in a company of 10,000 people.
- Imagine you randomly sample 5 people out of a company and they spend an amount of time in this activity as shown by the data points below.
- Is this statistically significant?
- Is it possible to estimate the chance the median time spent per person per day is between 15 and 40 minutes?



The Method of Measurement

Another Small Sample Example

THE *URN OF MYSTERY* PROBLEM

There is a warehouse full of thousands of urns.

Each urn is filled with over a million marbles, each of which are red or green.

The proportion of red marbles in each urn is unknown – it could be anything between 0% and 100% and all possibilities are equally likely.

Questions:
If you randomly select a single marble from a randomly selected urn, what is the chance it is red?
If the marble you draw is red, what is the chance the majority of marbles are red?
If you draw 8 marbles and all are green, what is the chance that the next one you draw will be red?

The Method of Measurement

Intuitions About Samples Are Wrong

- There are widely held misconceptions about probabilities and statistics – especially if they vaguely remember some college stats.
- These misconceptions lead many experts to believe they lack data for assessing uncertainties or they need some ideal amount before anything can be inferred.

"Our thesis is that people have strong intuitions about random sampling…these intuitions are wrong in fundamental respects...[and] are shared by naive subjects and by trained scientists"
Amos Tversky and Daniel Kahneman, Psychological Bulletin, 1971

The Method of Measurement

"Math-less" Statistics Table

Approximate 90% Confidence Interval	
Sample Size	N^{th} largest & smallest sample value
5	1^{st}
8	2^{nd}
11	3^{rd}
13	4^{th}
16	5^{th}
18	6^{th}
21	7^{th}
23	8^{th}

Simple Measurement Takeaway - This table makes estimating a 90% confidence interval of a population median easy.

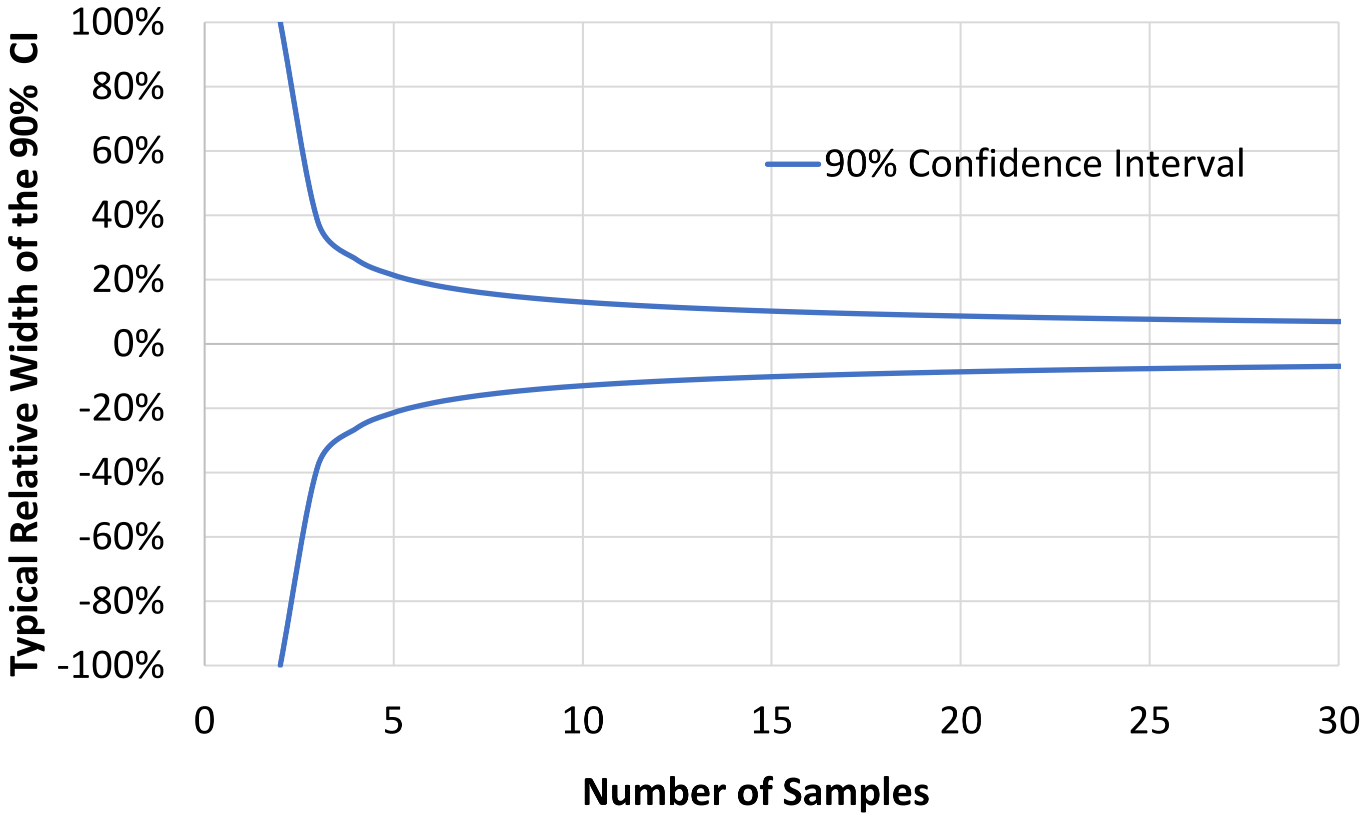
<u>The Rule of Five: </u>There is a 93.75% chance that the median of any population is between the smallest and largest values in a random sample of five.

This table expands on the Rule of Five. If you take 16 random samples of something, the 5^{th} largest and 5^{th} smallest values of that sample set approximate a 90% confidence interval.

The Method of Measurement

How Much Samples Can Tell Us

The graph below shows the average of relative reduction in uncertainty as sample sizes increase by showing the 90% CI getting narrower and narrower with each sample according to the student-t method.



With a few samples, there is still high uncertainty but…

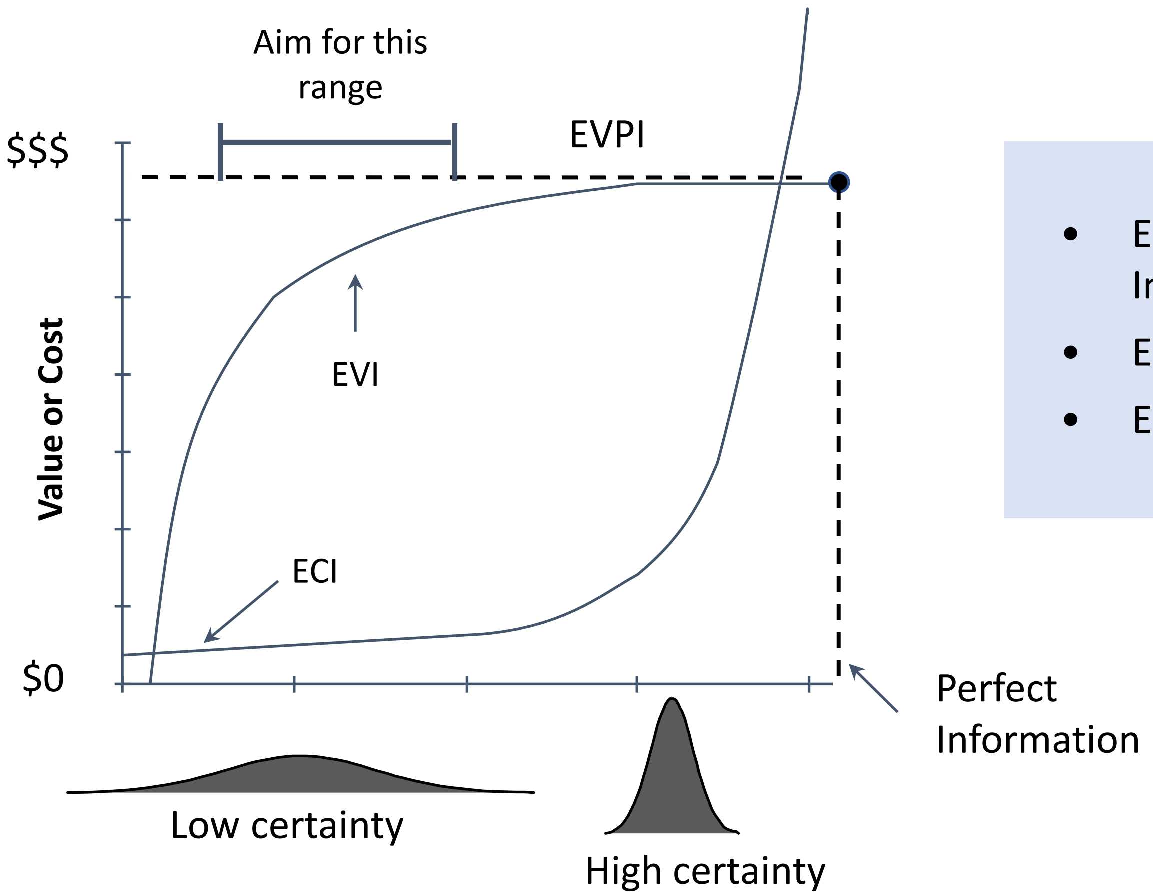
… each new sample reduces uncertainty a lot and the first few samples reduce uncertainty the most when initial uncertainty is high.

As number of samples increases, the 90 % CI get much narrower, but each new sample reduces uncertainty only slightly and beyond about 30 samples you need to quadruple the sample size to cut the error in half.

The Method of Measurement

The Value of Information

If we can model uncertainty about decisions, we can compute the value of information.



- EVPI – Expected Value of Perfect Information
- ECI – Expected Cost of Information
- EVI – Expected Value of Information

The Methods of Measurement

Three Useful Working Assumptions

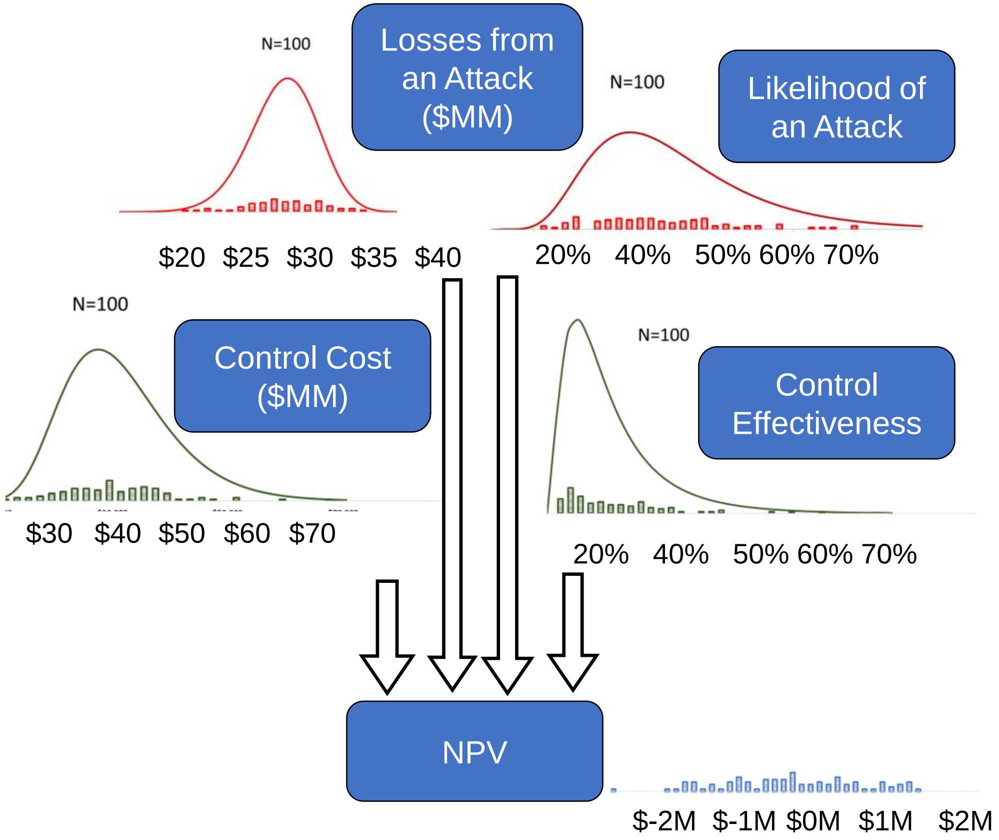
If your measurement is challenged with limited or messy data, consider the following:

- It's been measured before.
- You have more data than you think.
- You need less data than you think.

"It's amazing what you can see when you look"
Yogi Berra

The Method of Measurement

The Monte Carlo Simulation



Society of Petroleum Engineers (2000)

The Application of Probabilistic and Qualitative Methods to Asset Management Decision Making

G. S. Simpson, F. E. Lamb, J. H. Finch, and N. C. Dinnie

Abstract

Inter
comp
indus
the t

SSCAG/SCAF/EACE Joint International Conference (2008)

An Assessment of the Inherent Optimism in Early Conceptual Designs and Its Effect on Cost and Schedule Growth

D. Bearden, C. Freaner, R. Bitten, and D. Emmons

Abstract

When missions experience cost growth, cost estimators are often criticized for underestimating the cost of missions in the early conceptual design stage. The final spacecraft and instrument payload configuration at launch, however, can be significantly different as the project evolves, thereby leading to cost "growth" as

The Method of Measurement

Improving Expert Judgement

- Calibration of experts for overconfidence and inconsistency is a start.
- Decomposition tends to further improve expert estimates.
- We can leverage these facts for making improved models even without other recorded, empirical data (adding that comes next).

The Method of Measurement

Which Decomposition?

- Decomposition can improve models but not all decompositions are of equal value.
- Uninformative decompositions: Dwelling on speculations that you actually have no information about.
- Example: Assessing skill levels of unknown future attackers, speculating whether the risk is more the Russian mafia, Anonymous or China)

Simply improving the method of eliciting expert estimates is just a start

Now we need to inform the model with empirical data and continually update it based on new observations

The Method of Measurement

Informative Decompositions

Informative decompositions use what you know or data you can get to improve estimates in models.

Informative Decompositions:

- **Systems**: You have fairly detailed knowledge of your applications, what data they have and the hardware it runs on. Some of the parameters of these systems would change your estimate of a risk.
- **Types of Impacts**: You separate confidentiality, integrity and availability events. You have an idea of business volumes like sales and other processes. If a breach or outage occurred, you can describe something about the consequences.
- **Staff**: You have knowledge of the number of employees, device loss rates, and some knowledge of what data they may have.
- **Vendors & Customers**: You know who the parties you interact with and you have some knowledge about them.
- **Insurance:** Any cyber-insurance will have detailed language regarding limitations, exclusions, etc.

The Method of Measurement

Bayesian Methods

- "Bayesian" methods in statistics use new information to update prior knowledge.

Bayes Theorem:

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} = \frac{P(X)P(Y|X)}{\sum_i P(Y|X_i)\,P(X_i)}$$

$P(X)$ = the probability of X

$P(X|Y)$ = the probability of X given the condition Y

$\sum P(Y \mid X_i)\, P(X_i)$ = the sum of the probability of Y under each possible condition

- The Simplest Measurement Method — It turns out that calibrated people are already mostly "instinctively Bayesian".
  - Assess your initial subjective uncertainty with a calibrated probability
  - Gather and study new information
  - Give another subjective calibrated probability assessment

The Method of Measurement

The Rule of Succession

Danny Kahneman

A reference class is a population from which you draw observations of events to determine their frequency. Your "reference class" is much larger than you.

You can start by making as few assumptions as possible – your "baseline" uses only your reference class

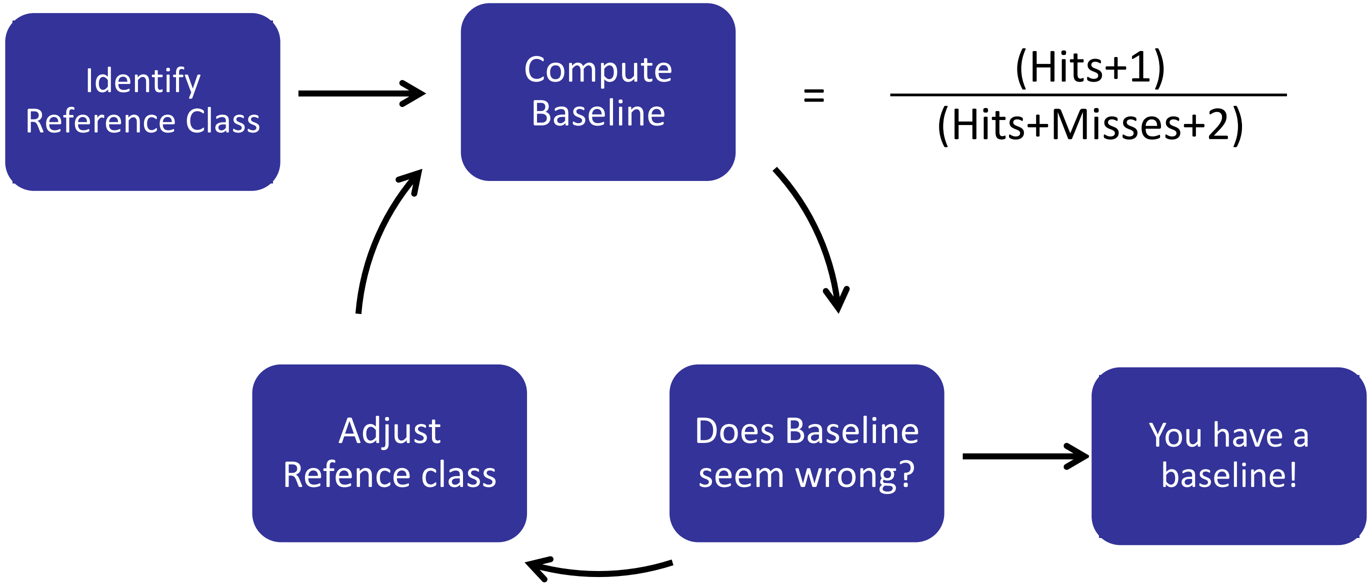
Pierre-Simon Laplace
1749-1827

- Laplace's "rule of succession": Given a population of reference class, like company-years, where some number of events occurred:
  - Chance of X (per year, per draw, etc.) =(1+hits)/(2+hits+misses)

The Method of Measurement

Computing Baseline Probabilities

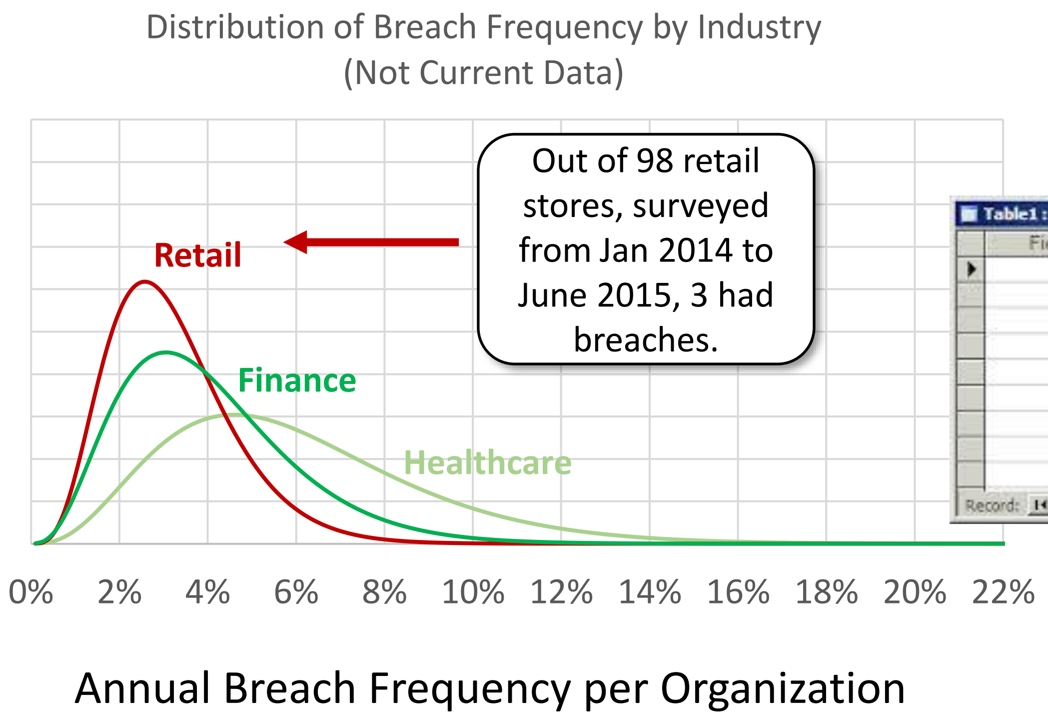
If the baseline seems too low or too high, it is probably because your reference class is larger than you first thought or because you believe a subset of it is more relevant.

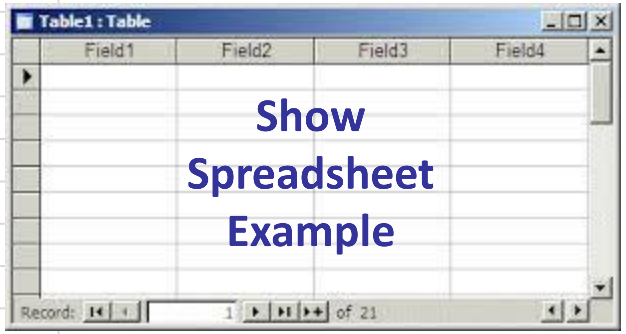


The Method of Measurement

Estimating Breach Rate w/History

- You have relatively few examples of major, reported breaches in each industry.
- There is a statistical method for estimating the frequency of breaches based on small samples.
- Spreadsheet for this at www.howtomeasureanything.com/cybersecurity





The Method of Measurement

Other Handy “Naïve Estimators”

Mean of a beta distribution is alpha/(alpha+beta).

alpha=observed hits +1, beta=observed misses+1

These are all the means of beta distributions to different questions. The alpha and beta are “hits and misses” but with one “free” hit and miss.

The chance of seeing an event that happened x times in y years in z organizations

=(1+x)/(2+yz)

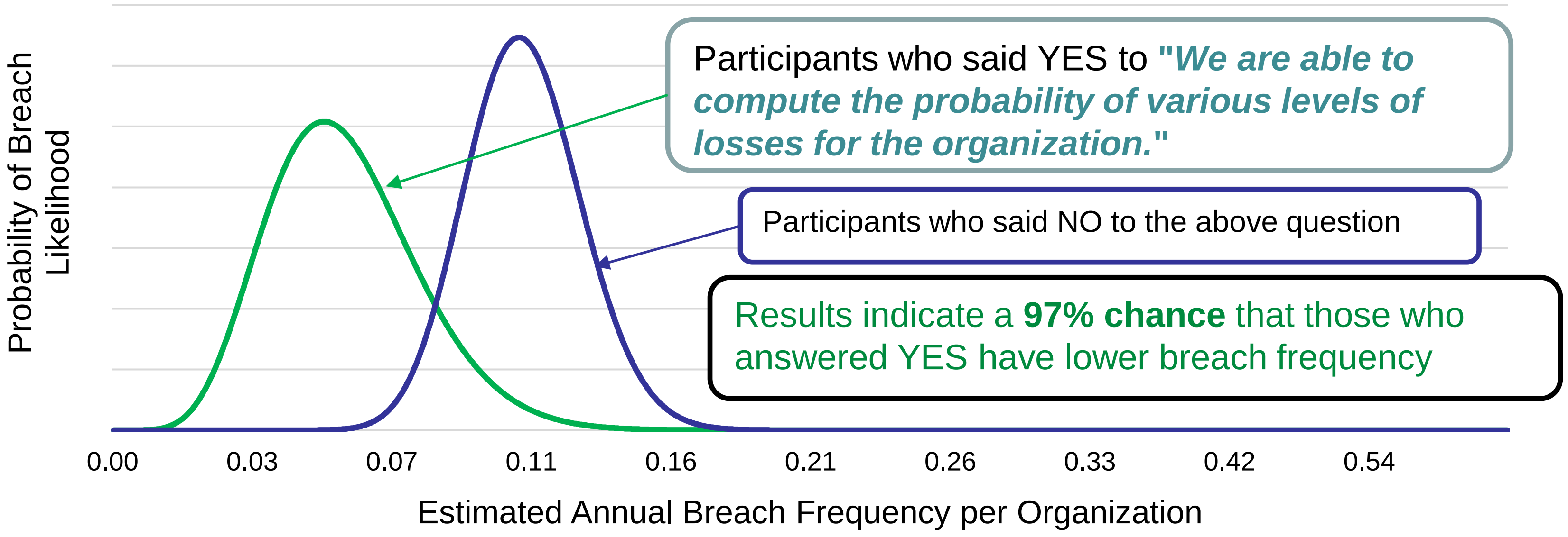
The chance that the next event will be worse than previous events:

=1/(1+n)

The Method of Measurement

What Reduces Data Breach Risk?

- The survey reveals another interesting result.
- Those who said they computed the probability of losses reported fewer breaches than those who did not.
- I would not treat this observation alone as sufficient – but it agrees with other independent evidence.



Summary

Final Thoughts

It's Been Measured Before	• Important topics have often been measured already.
You Have More Data Than You Think	• Define a reference class – don't commit the reference class fallacy.
You Need Less Data Than You Think	• Question your intuition about how and whether messy and incomplete data is.

Example Spreadsheets for many of the calculations mentioned can be found at www.howtomeasureanything.com

Do's and Don'ts

- Stop using risk matrices and "high, medium, low" as assessments of risk.

- Start using previously proven components:
    - probabilistic methods including Monte Carlo
    - calibrated experts
    - historical observations
    - quantified risk tolerance

Questions?

Contact:

Doug Hubbard

Hubbard Decision Research

dwhubbard@hubbardresearch.com

www.hubbardresearch.com

630 858 2788

Hubbard
Decision Research

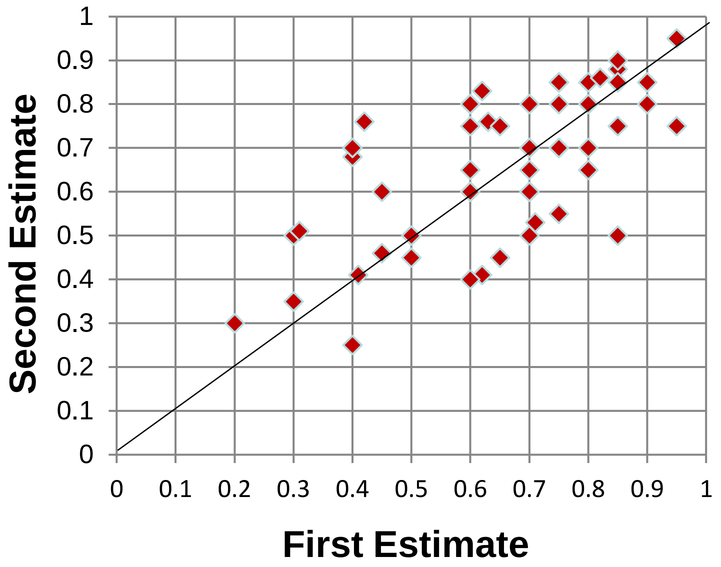
Supplementary Material

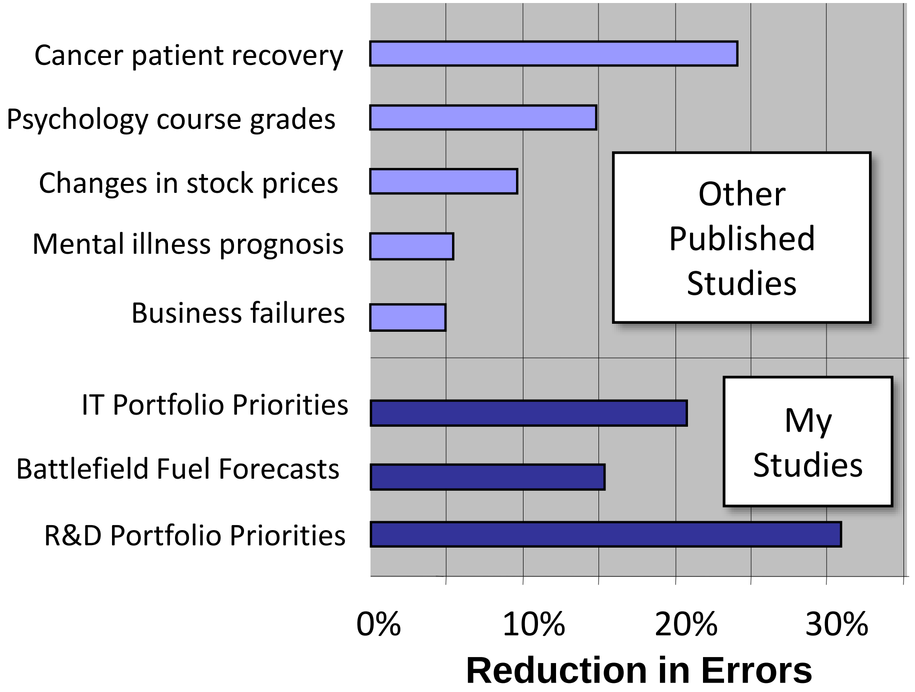
Hubbard Decision Research
2 South 410 Canterbury Ct
Glen Ellyn, Illinois 60137
www.hubbardresearch.com

The Method of Measurement

Measuring and Removing Inconsistency

Methods that statistically "smooth" estimates of experts show reduced error in several studies for many different kinds of problems.



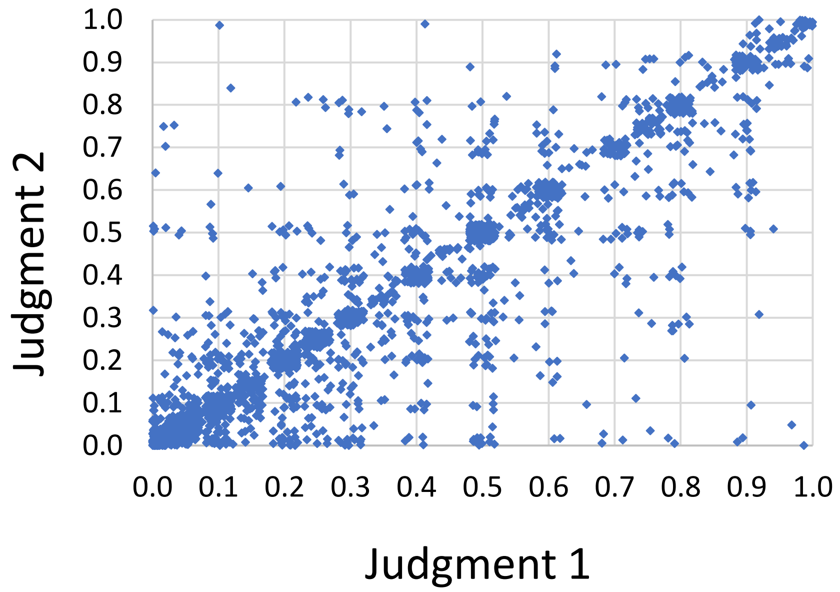


The Method of Measurement

Calibrating Expert Consistency

- We have gathered estimates of probabilities of various security events from:
  - 48 experts from 4 different industries.
  - Each expert was given descriptive data for over 100 systems.
  - For each system each expert estimated probabilities of six or more different types of security events.
- Total: Over 30,000 individual estimates of probabilities
- These estimates included over 2,000 duplicate scenarios pairs.

Comparison of 1st to 2nd Estimates of Cyber risk judgements by same SME



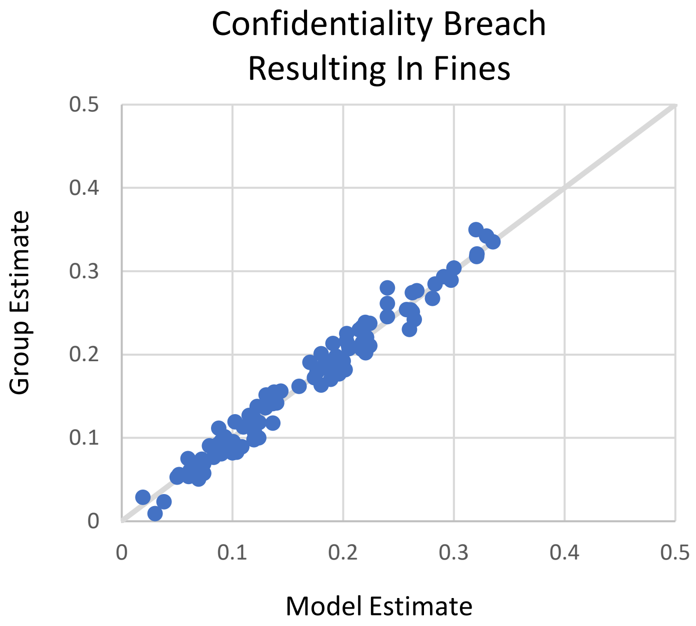
21% of variation in expert responses are explained by *inconsistency.*

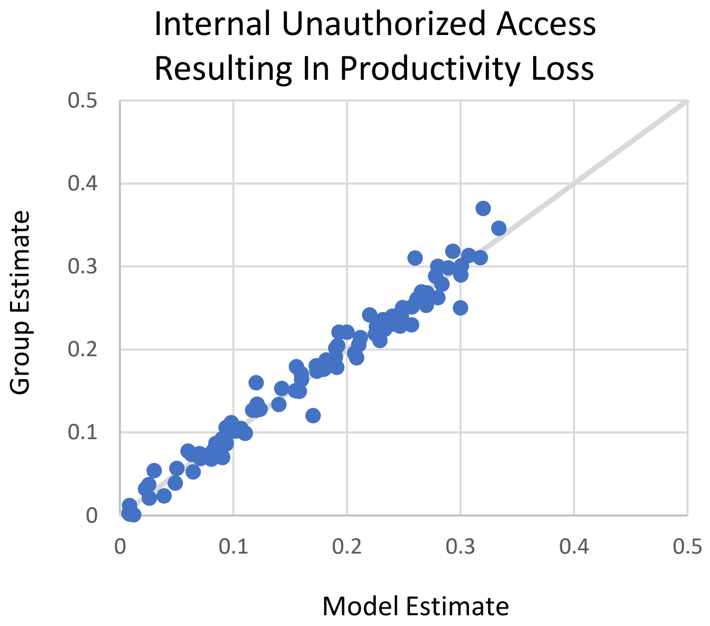
(79% are explained by the actual information they were given)

The Method of Measurement

Modeling Group Estimates of IT Security Event Likelihood

Examples of Models vs. Group Averages: Probabilities of different security events happening in the next 12 months for various systems prior to applying particular controls.





The models created produce results which closely match the group's average.

A large portion of the model error is due to judge inconsistency.

This nearly eliminates the inconsistency error.

The Method of Measurement

Logodds Model

A Logodds Model is a relatively simple approximation to “add up” a number of parameters that modify a probability when NPTs would be large.

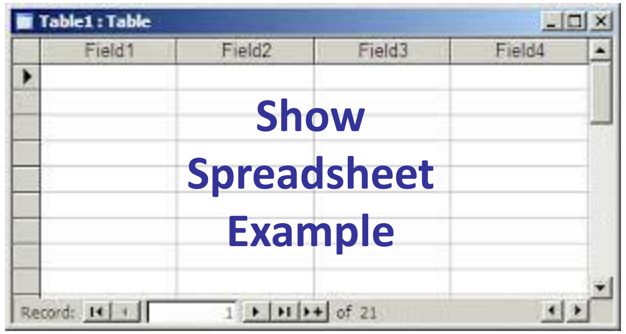
Logodds of X=LO(X)=ln(P(X)/(1-P(X))

Adjustment due to condition Y=A(Y) =LO(P(X|Y))– LO(P(X))

P(X|A,B,..)=Sum of (LO(A),LO(B),…)+LO(P(X)

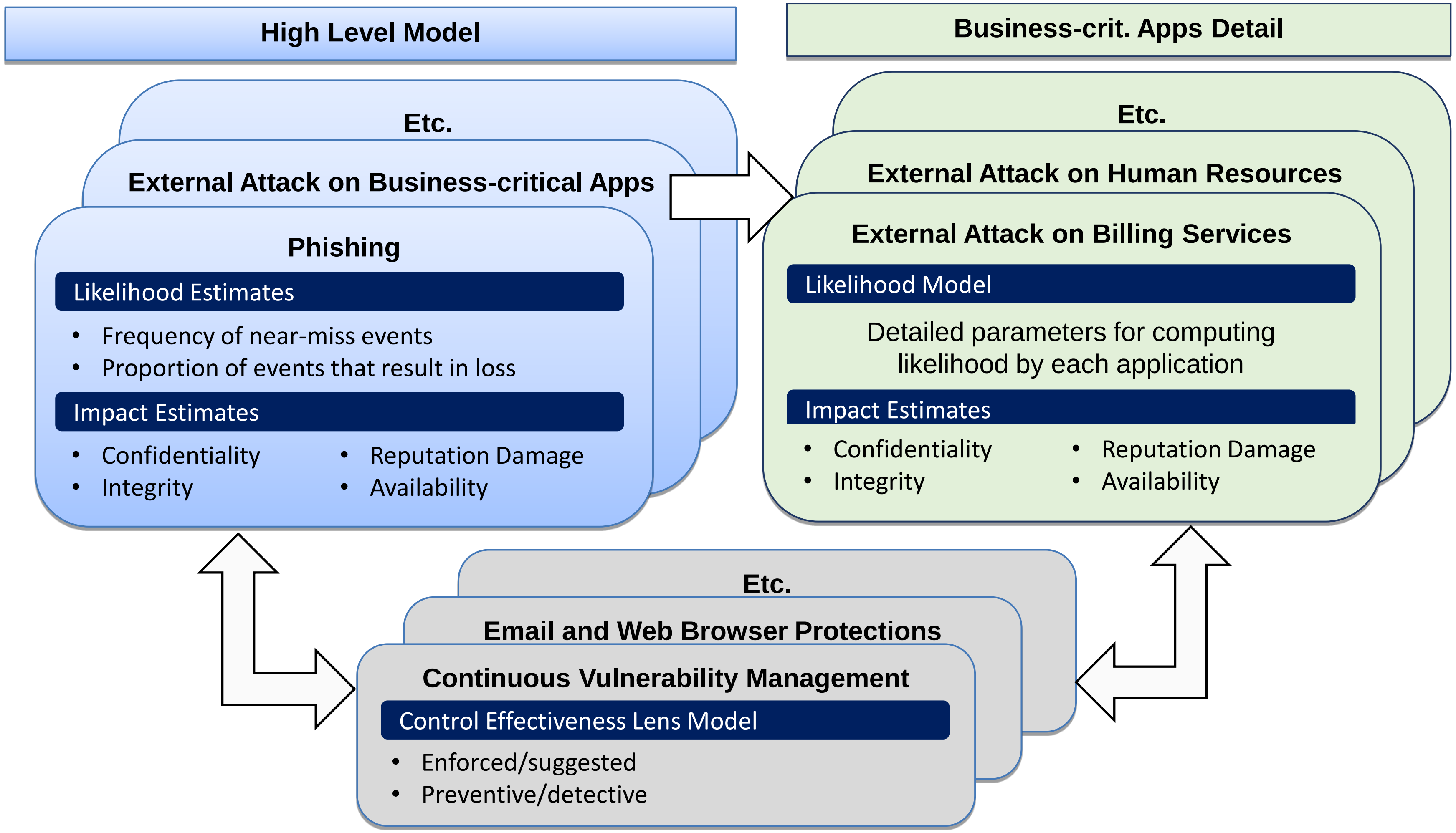
The more independent the parameter are, the better the Rasch approximation.

Initial Prob: P(E)	**10%**			
Baseline Logodds	-2.197			
	Conditions			
	A	**B**	**C**	**D**
P(E|X)	34.0%	15.0%	40.0%	12.0%
P(E|~X)	5.5%	9.0%	3.0%	8.0%
P(X)	16.0%	20.0%	19.0%	50.0%
Test P(E)	10.1%	10.2%	10.0%	10.0%
Logodds change|X	1.5339	0.4626	1.7918	0.2048
Logodds change|~X	-0.6466	-2.3136	-3.4761	-2.4423



The Method of Measurement

Cybersecurity Risk Model Structure



The Method of Measurement

Bayesian Methods: Node Probability Tables

Node Probability Table				
Condition				
A	B	C	D	P(E|A,B,C,D)
Yes	Yes	Yes	Yes	86%
No	Yes	Yes	Yes	40%
Yes	No	Yes	Yes	1%
No	No	Yes	Yes	2%
Yes	Yes	No	Yes	75%
No	Yes	No	Yes	40%
Yes	No	No	Yes	2%
No	No	No	Yes	1%
Yes	Yes	Yes	No	90%
No	Yes	Yes	No	35%
Yes	No	Yes	No	2%
No	No	Yes	No	1%
Yes	Yes	No	No	80%
No	Yes	No	No	40%
Yes	No	No	No	2%
No	No	No	No	2%

Conditional probabilities with combinations of conditions are recorded with an NPT.

With more than a few conditions and conditions that are more than binary, it will become unwieldly.

Recent models we created would have had thousands of rows.

The Method of Measurement

Lens vs. Logodds

The Lens method is more complicated to do well.

The Lens method is better at capturing more complex interactions between variables.

The Lens model will produce more realistic discrimination - logodds method tends to generate more extreme discrimination unless calibrated.

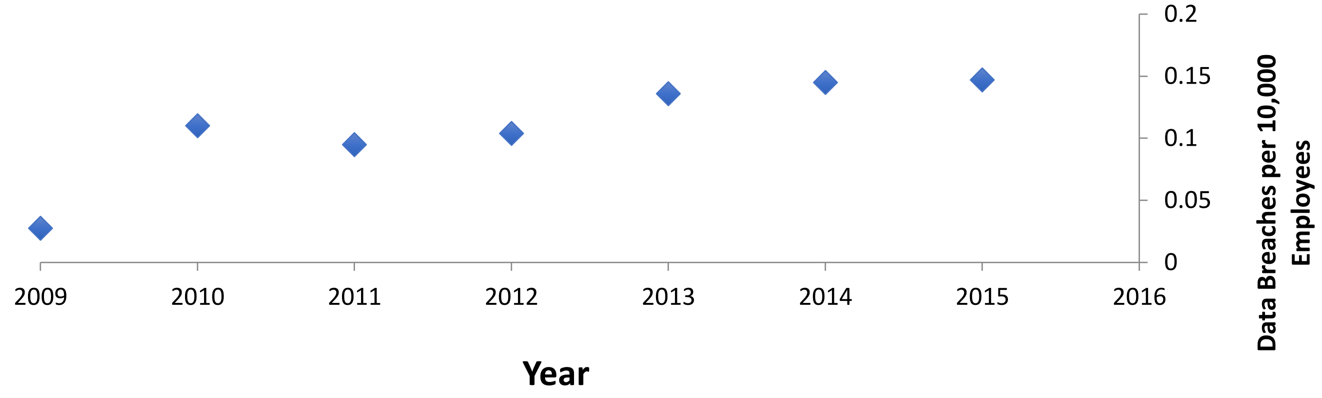
The Method of Measurement

Data Breaches/Yr vs. Number of Employees

- Data from the HHS “Wall of Shame” indicates that the rate of data breaches (more than 500 confidential records) is now consistently 14% per year per 10,000 employees.

- Estimating breach rate based on number of staff:

  P(breach|staff) = 1-(1-.14)^(staff/10000)

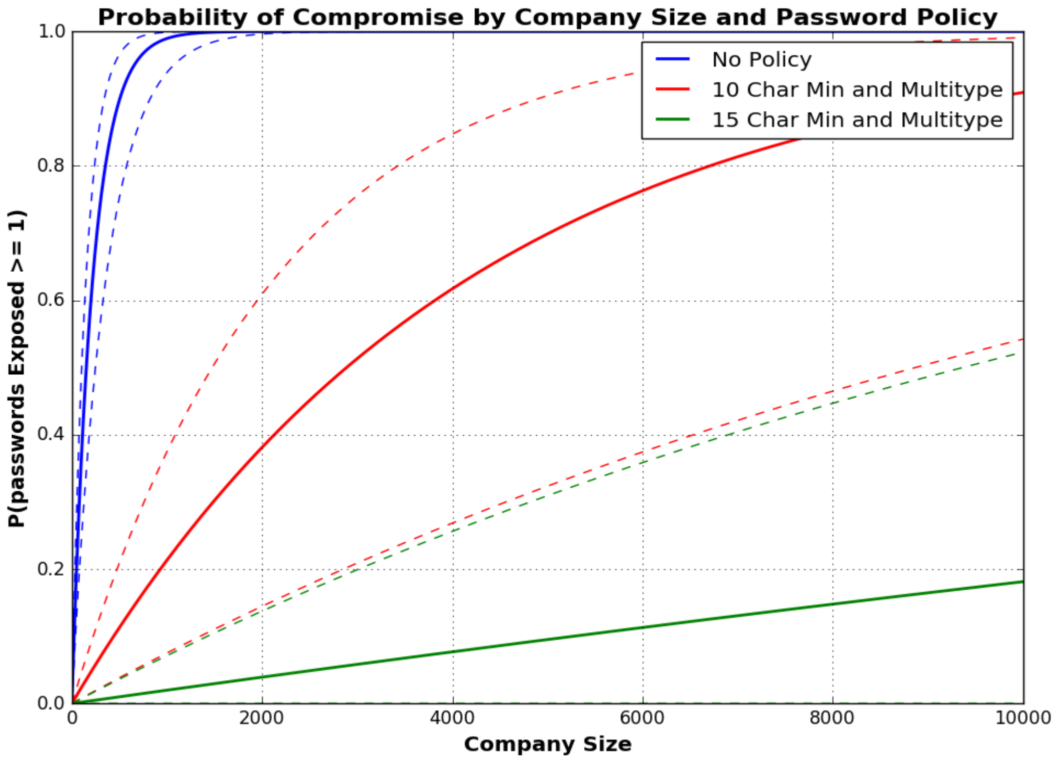


Source: Vivosecurity

The Method of Measurement

Password Compromise Statistics

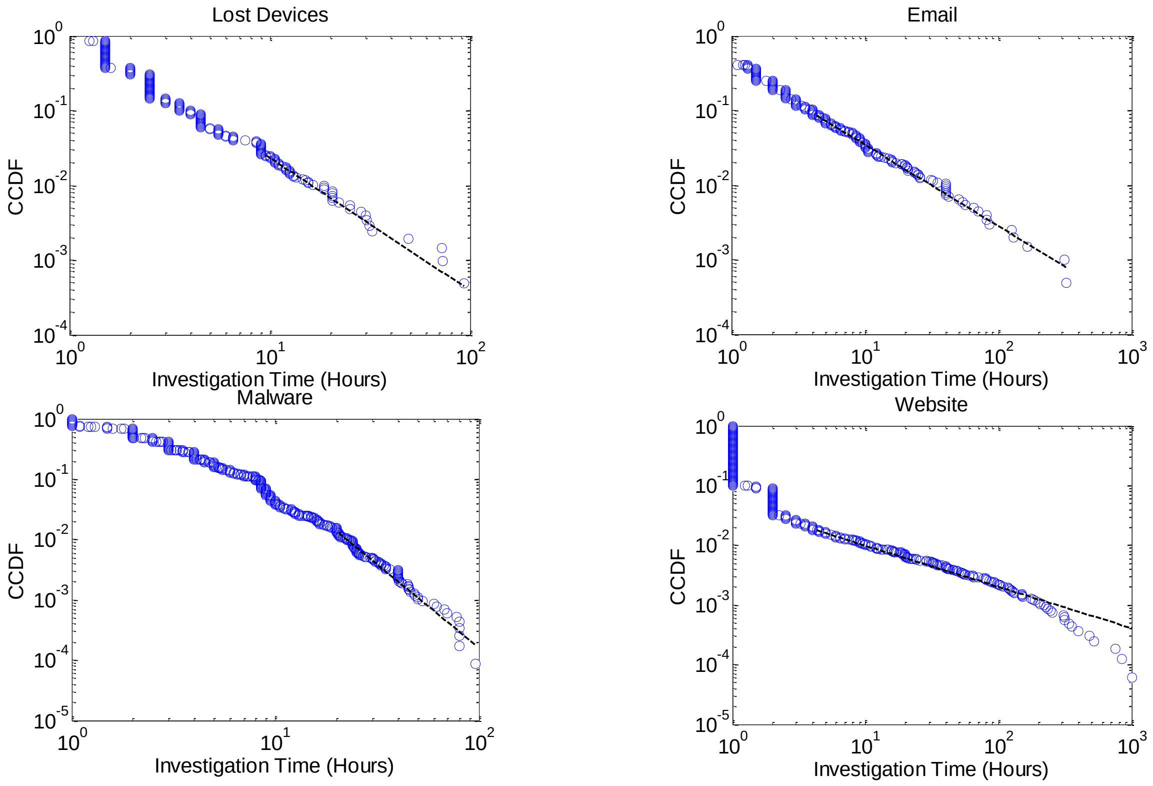
Source: Anton Mobley, GE Healthcare



The Method of Measurement

Power Laws in Investigation Times

The investigation times of several types of events are shown to have "Power Law" distributions. (Source: Marshal Kuypers)



The Method of Measurement

“Opinion Toward Quantitative Methods” (18 Questions)

18 questions on opinions of the use of quantitative methods in cybersecurity were asked. Here are some examples:

(Responses: Agree, Disagree, No Opinion/Don’t Know)

Information security is too complex to model with probabilistic methods.
Management and users won't understand the quantitative methods’ output.
An expert using quantitative probabilistic methods will do better risk assessments then an expert using intuition alone.

RESULTS: 80% of respondents had more “pro” than “anti” quantitative responses. Only 22% were consistently “pro” on quantitative and “anti” on softer scoring methods.

The Method of Measurement

The Stats Concepts Quiz (10 Questions)

EXAMPLE: Assume that you have a portfolio of systems for which you have observed no security events in the past year that resulted in a monetary or productivity loss, which of the following statements is true?

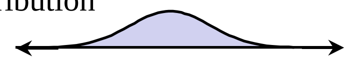
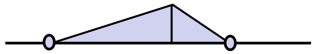
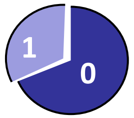
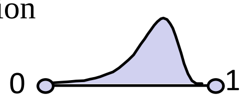
Answer Options	Response Percent
If no events were observed, then we have no data about the likelihood of these events.	2.2%
The fact that no events were observed tells us something about the likelihood of these events.	37.0%
One year is not long enough time to gather enough observations to make an inference.	4.4%
Since some events may not have been observed, the lack of observed losses tells us nothing.	31.9%
There is insufficient information to answer the question.	17.8%
I don't know	6.7%

The Method of Measurement

Basic Distributions

Each of these examples can be found on

www.howtomeasureanything.com/cybersecurity

Distributions*	Upper & Lower Bound	Best Estimate
Normal distribution 	Represents the "90% confidence interval"	Always half-way between upper and lower bound
Lognormal distribution 	Represents the "90% confidence interval"; the absolute lower bound of a lognormal is always 0	Always a function of the upper and lower bound
Uniform distribution 	Represents the absolute (100% certain) upper and lower bounds	NA
Triangular distribution 	Represents the absolute (100% certain) upper and lower bounds	Represents the mode; the most likely value
Binary distribution 	NA	Represents the % chance of the event occurring
Beta distribution 	Generates a value between 0 and 1 based on “hits” and “misses”	The mode of a beta is (hits-1)/(hits+misses-2)

*A “●” means a “hard” stop, an “➔” arrow means unbounded

Selected Sources

Tsai C., Klayman J., Hastie R. "Effects of amount of information on judgment accuracy and confidence" *Org. Behavior and Human Decision Processes,* Vol. 107, No. 2, 2008, pp 97-105

Heath C., Gonzalez R. "Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making" *Organizational Behavior and Human Decision Processes,* Vol. 61, No. 3, 1995, pp 305-326

Andreassen, P." Judgmental extrapolation and market overreaction: On the use and disuse of news" *Journal of Behavioral Decision Making*, vol. 3 iss. 3, pp 153-174, Jul/Sep 1990

Williams M. Dennis A., Stam A., Aronson J. "The impact of DSS use and information load on errors and decision quality" *European Journal of Operational Research,* Vol. 176, No. 1, 2007, pp 468-81

Knutson et. al. "Nucleus accumbens activation mediates the influence of reward cues on financial risk taking" *NeuroRepor*t, 26 March 2008 - Volume 19 - Issue 5 - pp 509-513

A small study presented at Cognitive Neuroscience Society meeting in 2009 by a grad student at U. of Michigan showed that simply being briefly exposed to smiling faces makes people more risk tolerant in betting games.

Risk preferences show a strong correlation to testosterone levels – which change daily (Sapienza, Zingales, Maestripieri, 2009).

Recalling past events that involved fear and anger change the perception of risk (Lerner, Keltner, 2001).