

# The Failure of Risk Management



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













#### 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



- The Meta-Decision
- Getting Started
- **Obstacles** •
- Simple Math





### Introduction

#### A Few Events from the Last 10 Years

- Fukushima Daiichi nuclear disaster (2011)
- Deepwater Horizon offshore oil spill (2010)
- Flint Michigan water system (2012 to present)
- Samsung Galaxy Note 7 (2016)
- Multiple large data breaches (Equifax, Anthem, Target)
- Amtrak derailments/collisions (2018)
- California utility PG&E wildfires (2018)
- COVID (2020)













### **Question: What is your single biggest risk?**

Answer: How you measure risk.



#### Introduction

#### Types of Measurement Methods





Share of Methods Used in Cybersecurity Risk Assessment





#### Do "Scores" and "Scales" Work?

#### The Ubiquitous Risk Matrix





### 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?





Bad

1 2 3 4

### Do "Scores" and "Scales" Work?

Unintended consequences of simple scoring methods





Good

5



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







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.



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

#### <u>Abstract</u>

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



<u>A</u>

A



#### 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.



### 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."

PAUL E. MEEHL CLINICAL VERSUS STATISTICAL PREDICTION

Philip Tetlock tracked a total of over 82,000 forecasts from 284 experts in a 20year 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





Monte Carlo: How to Model Uncertainty in Decisions



#### What Published Research Says (See sources slide for details)

- Psychologists showed that simple decomposition greatly reduces estimation error for estimating the most uncertain variables.
- In the oil industry there is a correlation between the use of quantitative risk analysis methods and financial performance.
- Data at NASA from over 100 space missions showed that Monte Carlo simulations and historical data beat softer methods for estimating cost and schedule risks.



#### Why Does Our Risk Tolerance Change?





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

#### https://www.howtomeasureanything.com/riskmanagement/

Event	Event Probability	Impact (90% Confidence Interval)		Random Result (zero when the
	(per Year)	Lower Bound	Upper Bound	event did not occur)
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
₽	Ŷ	$\mathbf{\nabla}$	Ŷ	$\mathbf{\nabla}$
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.

Ta	ble1 : Table				
	Field1	Field2	Field3	Field4	-
		She	אור		
		0			-
		Spread	Isheet		
		Exan	nple		
			220	11.576	-
Recor	d: 14 4	1 + + +	+ of 21	1	•



### Obstacles to Better Methods

Obstacles: Why Better Methods Are Not Adopted





### So Why Don't We Use More Quantitative Methods?

Commonly stated reasons for not using quantitative methods



## 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



© Hubbard Decision Research, 2020



- "Experience is inevitable, learning is not." Paul Schoemaker
- Kahneman and Klein differentiate high and low validity tasks based on feedback:



- Consistent
- Immediate
- Unambiguous





### Irrational Bias Against Algorithms

A Double Standard



Exsupero Ursus

A common form of the *Exsupero Ursus* fallacy: "The quantitative model must have

- All the variables 1)
- 2) All the data
- 3) All the right distributions and correlations
- All the above 4)

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			



What Measurement Really Means





#### 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.





### Calibrated Experts

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.



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.





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.





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 American Psychological Associ
2015, Vol. 21, No. 1, 1–14	1076-898X/15/\$12.00 http://dx.doi.org/10.1037/xap000
The Psychology of Intelligence An	alysis: Drivers of Prediction
Accuracy in Wor	ld 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 Tetlo	ck
University of Penns	yIvania
This article extends psychological methods and concepts in	to a domain that is as profoundly consequen-
tial as it is poorly understood: intelligence analysis. We re	port findings from a geopolitical forecasting
tournament that assessed the accuracy of more than 150,00	0 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 perfor	mance 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	
OBJECT of Measurement	The thing being measured is not well defined.



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.



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 Three Misconceptions Behind Any Perceived "Immeasurable"

#### The Method of Measurement

METHOD of Measurement	Many procedures of empirical observation are misunderstood.



#### The Urn of Mystery



#### THE URN OF MYSTERY PROBLEM

- There is a warehouse full of urns.
- Each urn is filled with over a <u>million</u> 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?



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 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)



Estimating Breach Rate w/ "Reference Classes"

- You may not have had a particular major event type, but others have. •
- You have relatively few examples of major, reported breaches in each industry. •
- There is a statistical method for estimating the frequency of events based on small samples.



Annual Breach Frequency per Organization



Bayesian Methods

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



- 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



Final Thoughts

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

# Example Spreadsheets for many of the calculations mentioned can be found at <u>www.howtomeasureanything.com</u>



Your Real Job in Risk Assessment & Management

You are a creator and manager of models – not just a "down in the weeds" estimator/forecaster/decision maker.

- You evaluate data from external literature and reference classes.
- You frequently record internal estimates and decisions, whether large or small.
- You evaluate performance, continuously improve, and look for the best forecasters.
- This holds for models of expert intuition (including your own) and complex calculations.
- You gradually replace areas of pure intuition with tested calculations.



- 1. How are measurement instruments (including experts) calibrated?
- 2. How are probabilities updated with empirical data?
- 3. How are probabilities and impacts modeled/aggregated?
- 4. How are resource allocation decisions made toward mitigating risks?
- 5. How is the performance of method itself being measured and updated?
- 6. How is completeness and correctness verified?
- 7. How do we implement it?





• 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



Contact:

Doug Hubbard

Hubbard Decision Research

dwhubbard@hubbardresearch.com

www.hubbardresearch.com

630 858 2788



# Supplementary Material

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



### **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



- 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" NeuroReport, 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).