



Innovations in the Science and Practice of Decision Analysis: The Role of *Management Science*

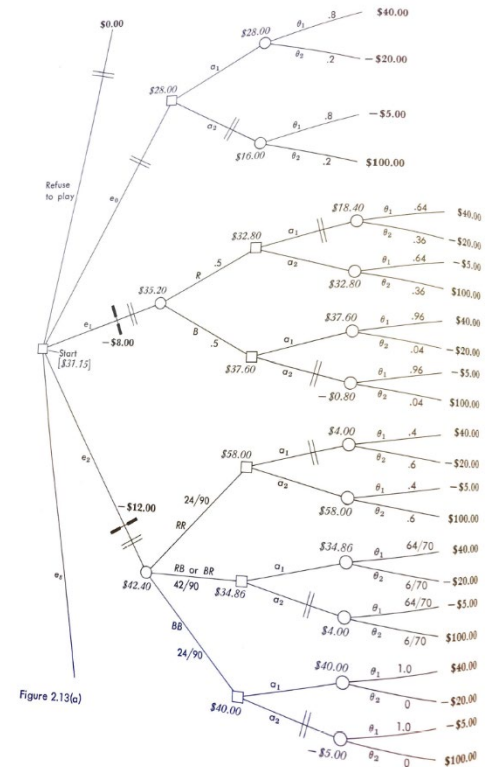
Reprise of
Management Science's
65th Anniversary Conference
Boston University
Questrom School of Business
May, 2019

April, 2022

Jim Dyer and Jim Smith

What is Decision Analysis (DA)?

- For each alternative:
 - Assess utilities for possible outcomes
 - How to calculate or forecast values?
 - How to assess utility for money?
 - How to make tradeoffs among objectives?
 - Assess probabilities: **Descriptive** comes
 - Common mistakes?
 - How to assess probabilities?
 - How to combine probabilities **Prescriptive** es?
 - Multiply utilities and probabilities and sum to find the expected utility for the alternative
- Choose alternative with maximum expected utility
- Procedure dates back to Bernoulli (1738) and was formally justified by Savage **Normative** Neumann-Morgenstern (1944/7), DeFinetti (1937), Ramsey (1931), ...



From Raiffa (1968)

DA has a long history in MS

- Flood (1955 in MS), Decision Making:

“L. J. Savage, in his *Foundations of Statistics*, offers a probability-utility type theory of decision that shows the close logical connection between any such theory and a very few plausible assumptions about rational behavior. In fact, if the over-all normative problem is in some sense necessarily one requiring probabilistic considerations of valuations leading to conscious choices among known classes of alternatives, then it seems likely that a good many of these interestingly complex mathematical findings will have practical importance.”

Flood also cited Edwards (1954) “Theory of Decision Making”

- Allais (1957 in MS, 53 pages!):

METHOD OF APPRAISING ECONOMIC PROSPECTS
OF MINING EXPLORATION OVER LARGE
TERRITORIES

Algerian Sahara Case Study*

M. ALLAIS†

Summary

1. The present study has essentially been conceived as a part of a larger study in the field of operations research. Its purpose resides in the scientific detection of the best and economically optimal strategy to be used in prospecting for metal deposits in the Sahara.

It is designed to provide a reasonable forecast of the economic prospects of mining exploration in this region, both in the light of statistical distributions of expected deposits and of their distribution according to value.

Won the 1957 Lanchester Prize

DA has a long history in MS (II)

- Some early MS papers:

Borch (1963), A
Note on Utility and
Attitudes to Risk

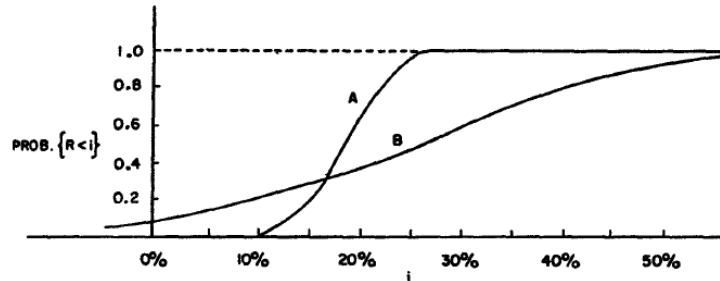


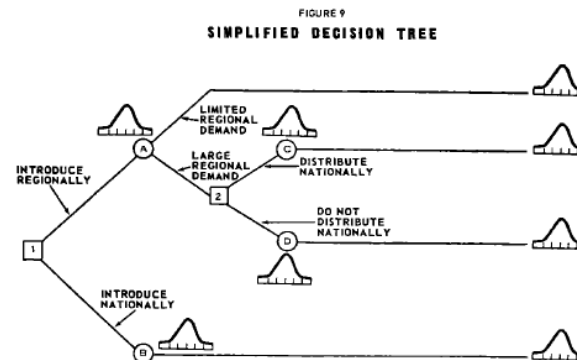
FIG. 1. Comparison of the cumulative distribution functions of R for the investments in model A and in model B.

Hillier (1963), The Derivation of Probabilistic
Information for the Evaluation of Risky Investments

6. The assumption in paragraph 1 implies that we must have

$$f(E, V) = \int_{-\infty}^{+\infty} u(x) dF(x).$$

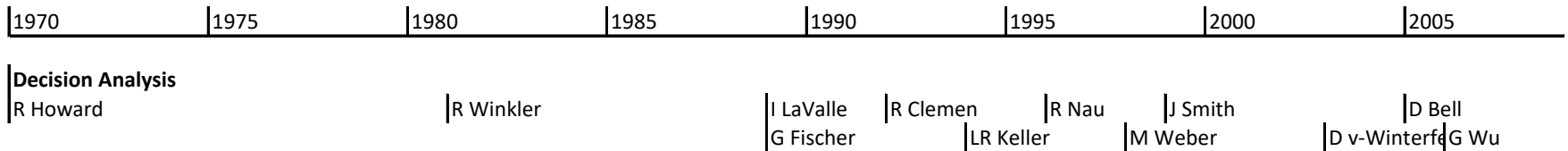
If $F(x)$ is an arbitrary probability distribution, the right-hand side can be a function of E and V alone only if $u(x)$ is a polynomial of second degree.



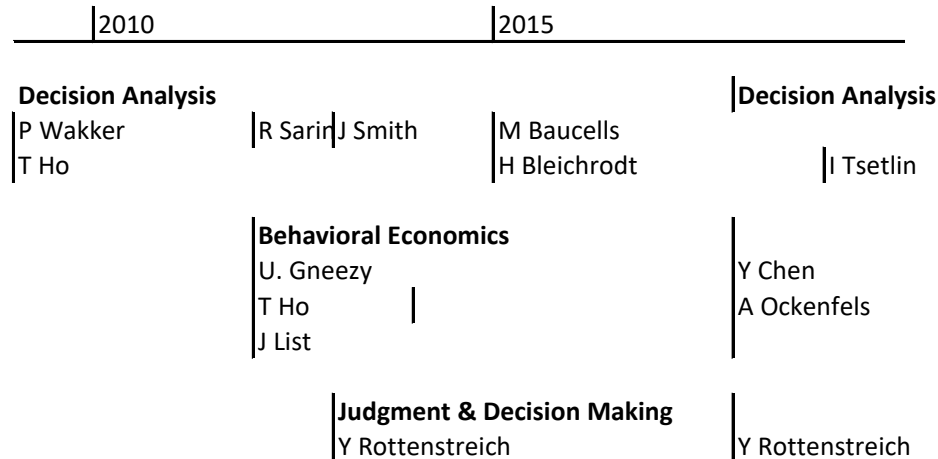
Hespos, Strassman (1965), Stochastic Decision
Trees for the Analysis of Investment Decisions

DA department history at MS:

- The DA department started in 1970 and focused on theory/ methodology; added a behaviorally oriented DE in 1989:



- MS added BE and JDM departments in 2011 and 2012; recombined into DA department in 2018:



Now called
“Behavioral
Economics and
Decision Analysis”

Added Marie Claire Villeval
as another DE

Observation:

- MS has been a key outlet for scientific research in DA with a focus on theory and methodology.
- Applications are typically published elsewhere, if at all:
 - *Operations Research (OR), Interfaces, Decision Analysis*
 - Field journals in medicine, environment, petroleum engineering, ...
 - Consulting or corporate applications are rarely published
- For MS, this is as was intended. Churchman (1994):

“My hope was that MS would be quite different from OR, because MS, the journal, the meetings, and the research would be the attempt to create and design a science of management that lived up to the standards of good science, whereas OR would be the practical application of that science.”

Challenge:

- How to trace the impact of research published in *MS* on practice?
 - *MS*'s focus on theory and methodology without publishing applications makes it difficult to identify the direct impact of DA research on practice.
 - We believe the impact is real and significant but one has to “pull the thread” to reveal the connections and influence.



WHAT IS AN APPLICATION AND WHEN IS THEORY A WASTE OF TIME?*

M. SHUBIK

Martin Shubik's classification of game theories:

- High church: “Mathematics, axioms, and formal solution concepts”
 - “Much of it can verge on ‘art for art’s sake.’”
 - Research moves “one step closer to operating concerns but without direct or immediate application.”
- Low church: Involves “work on a specific application”
 - “Produces actual calculations, if only for illustrative purposes, and possibly parametric sensitivity analysis”
 - “Of some, but nevertheless relatively modest, worth, but nowhere near the applied value of linear programming.”
- Conversational: “Advice and suggestions about thinking strategically”
 - Examples include understanding zero-sum games, nonzero-sum games, the prisoner’s dilemma, ...
 - “Of considerable worth.”

Plan for the rest of talk:

- Research published in *MS* is typically high church; academics are accused of “talking to each other.”
- We will:
 - Focus on some high-church theory and methodology research that has appeared in *MS* and
 - Talk about how it has impacted (or could impact) low-church practice and conversation
- Two topics:
 - Multiattribute utility theory (Jim Dyer)
 - Probability assessment (Jim Smith)

Early work on multiple criteria in MS

- Several ad hoc approaches to making decisions with multiple objectives appeared in the early 1960s
 - Terry (1963) Comparative evaluation of performance using multiple criteria
 - Eckenrode (1965) Weighting multiple criteria

	2.TROOP SAFETY	3.MOBILITY	4.ECONOMY	5.RELIABILITY	6.LETHALITY
1. EARLY AVAILABILITY					
2. TROOP SAFETY					
3. MOBILITY					
4. ECONOMY					
5. RELIABILITY					

FIG. 2. Partial paired comparisons I: J puts the number of the more valuable criterion of the pair being considered in each block. Example: 2 in the "early availability—troop safety" block means J judges the latter to be more valuable for the system in question.

Motivation: Programming, Planning, and Budgeting systems for the military

- Black (1964) Systems analysis in government operations

The Relation of the Systems Approach to Program Budgeting

PPBS and the Systems Approach

Program Packaging and Budgeting (PPBS) is a method of financial planning and control of government operations. It refers to the structuring of governmental activities into packages, composed of elements which can rationally be

The beginnings of a formal theory for Multiattribute Utility (MAU)

- High-church Theory:
 - An early extension of single attribute utility to multiple attributes was provided by Debreu in 1959 with other advances by Krantz, Luce, and Tukey and many others in the 1960s.
 - Fishburn (1968, MS) *Utility Theory* is a summary of this research and contains a proof of the following proposition:

Proposition 5.1 (Additive Utilities). A number $u_i(x_i)$ can be assigned to each x_i in X_i , for $i = 1, 2, \dots, n$, so that if $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ are in X then

$$(5.1) \quad x \preceq y \quad \text{if and only if} \quad u_1(x_1) + u_2(x_2) + \dots + u_n(x_n) \\ \leq u_1(y_1) + u_2(y_2) + \dots + u_n(y_n).$$

RAND research on MAU in the 1960s

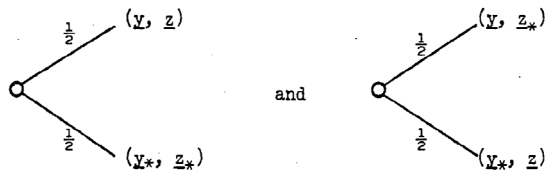
- Low church:

- K. R. MacCrimmon (1968), Decisionmaking Among Multi-Attribute Alternatives: A Survey and Consolidated Approach, RM-823-ARPA
- James R. Miller (1969), Assessing Alternative Transportation Systems, RM-5865-DOT

- High church:

- Howard Raiffa (1969), Preferences for Multi-Attributed Alternatives, RM-5868-DOT

Marginality Assumption: For any y, z , he feels indifferent between the following two lotteries:



Substituting $(2a, b, c)$ into (1), we get the additive representation

$$u(y, z) = u_Y(y) + u_Z(z) . \quad (2d)$$

This result, although exposted differently here, is due to Fishburn.

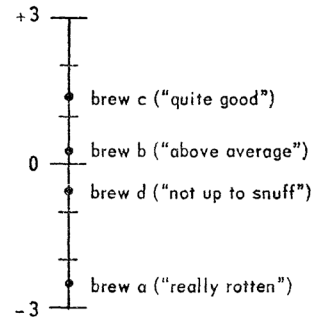
Now let's see how we can use $(2a, b, c)$ in obtaining measurements.

MAU Theory in MS

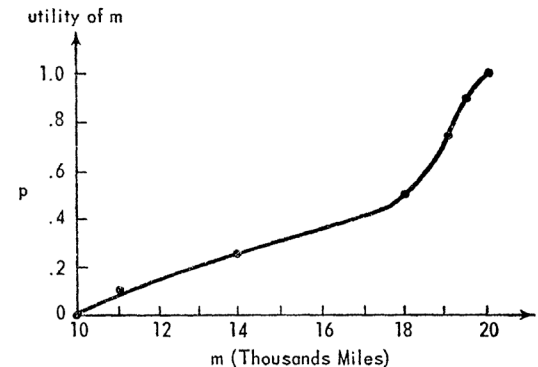
- MS articles in the spotlight

- Peter Fishburn (1967),
Methods of Estimating
Additive Utilities

Reviews 24 methods for
estimating additive utility functions



a. Method 2



b. Method 4

- Ralph L. Keeney (1972),
Utility Functions for Multi-
attributed Consequences

Operational assumptions are postulated
about a decision-maker's preferences and
functional forms of utility functions satisfying
those forms are derived

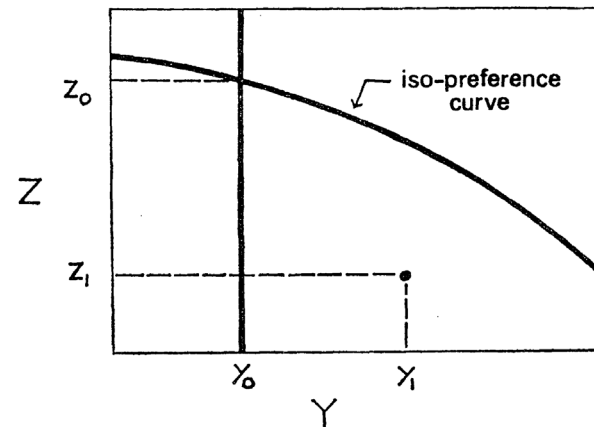
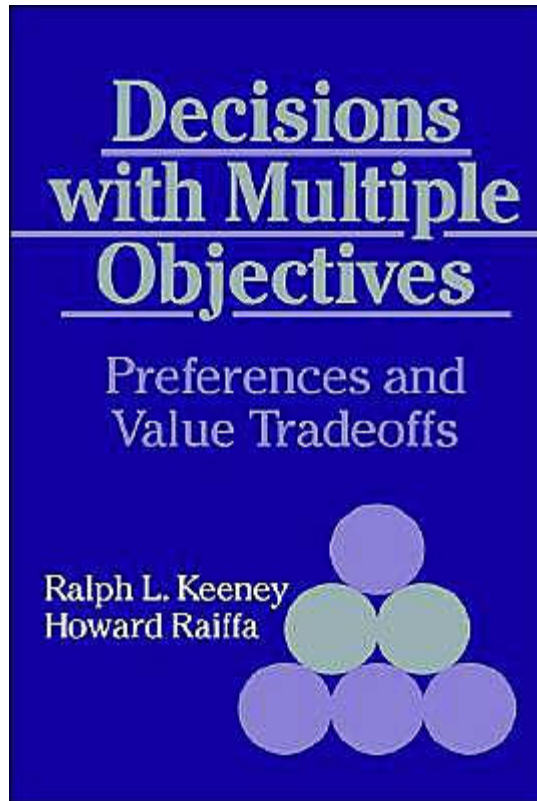


FIGURE 2

MAU made simple

- The collaboration between Keeney and Raiffa led to the classic book on MAU originally published in 1976



High church: Originally published in 1976, this book provides an accessible summary of MAU theory drawing on the work of Fishburn and others along with original contributions by Keeney and Raiffa.

Low church: The book also contains several chapters on practical methods for assessing utility functions, weights on objectives and for creating objectives hierarchies.

Recognition: Co-winner of the 1976 Lanchester Prize

Applications in MS after K&R

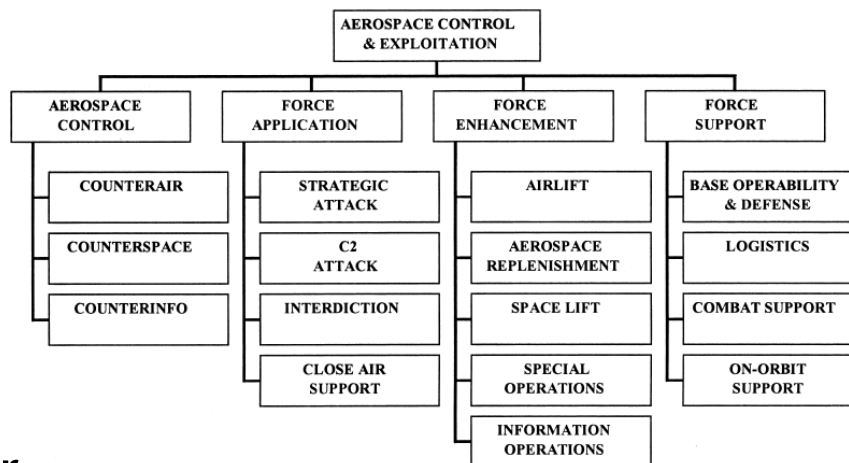
Public Sector Applications

- Bodily (1978) Police Sector Design Incorporating Preferences of Interest Groups for Equality and Efficiency
- Crawford, Hutzinger, Kirkwood (1978) Multiobjective Decision Analysis for Transmission Corridor Selection
- Ford, Keeney, Kirkwood (1978) Evaluating Methodologies: A Procedure and Application to Nuclear Power Plant Siting Methodologies
- Golabi, Kirkwood, Sicherman (1981) Selecting a Portfolio of Solar Energy Projects using Multiattribute Preference Theory
- Keeney, Sarin, Winkler (1984) Analysis of Alternative National Ambient Carbon Monoxide Standards
- Keeney, von Winterfeldt, Eppel (1990) Eliciting Public Values for Complex Policy Decisions
- Gregory, Keeney (1994) Creating Policy Alternatives Using Stakeholder Values
- Grushka-Cockayne, de Reyck, Degraeve (2008) An Integrated Approach for Improving European Air Traffic Management

Other applications in MS

Military

- Stafira, Parnell, Moore (1997) A Methodology for Evaluating Military Systems in a Counterproliferation Role
- Parnell, Conley, Jackson, Lehmkuhl, Andrew (1998) Foundations 2025: A Value Model for Evaluating Future Air and Space Forces



But is it a **measurable** multiattribute value model?

See Smith and Dyer “On (Measurable) Multiattribute Value Functions: An Expository Argument,” *Decision Analysis*, 2021

Private Sector

- Keefer (1991) Resource Allocation Models with Risk Aversion and Probabilistic Dependence in Offshore Oil and Gas Bidding

Impact of MAU on other societies



- Optimization with multiple criteria
 - Charnes, Cooper, Ferguson (1955, MS)
 - First paper on goal programming which can be viewed as using a piecewise linear approximation to a multi-attribute utility function as the objective function.
 - Geoffrion, Dyer, Feinberg (1972, MS)
 - First paper on interactive multi-criterion optimization based on MAU
 - Zionts, Wallenius (1976, MS) and others
 - Interactive Algorithms for solving MCDM problems
- Subsequent work led to the formation of the International Society on Multiple Criteria Decision Making and the INFORMS Multiple Criterion Decision Making Section

	A	B	
COST	2150	2100	
HP	108.3	108.3	
MPG	29	27	
WHICH DO YOU PREFER? IF YOU ARE INDIFFERENT, TYPE I			
A			
	A	B	
COST	2150	2100	
HP	108.3	108.3	
MPG	29	28	
WHICH DO YOU PREFER? IF YOU ARE INDIFFERENT, TYPE I			
B			
	A	B	
COST	2150	2100	
HP	108.3	108.3	
MPG	29	27.5	
WHICH DO YOU PREFER? IF YOU ARE INDIFFERENT, TYPE I			
I			
THE TRADEOFFS ARE			
COST	-1		
HP	10		
MPG	33.33		
NEW OPERATING POINT			
2500 140 20.5			
NEW DECISION VECTOR			

Estimation of w_s^k

The vector w^k

Computed by Frank-Wolfe Algorithm (9)

Impact of MAU on Medical Decision Making

Early work in MS:

- Stimson (1969) Utility Measurement in Public Health Decision Making
- Torrance (1976) Health Status Index Models: A Unified Mathematical View

Many other articles published on DA applications in health care in the 70s and 80s

Quality-Adjusted Life Years (QALYs)

- Fanshel, Bush (1970, OR) A Health Status Index and its Application to Health-Services Outcomes

One of the first articles to reference utility theory as the basis for a health status index:

- Plishkin, Shepard, Weinstein (1980, OR) Utility Functions for Health and Life Years
Uses independence assumptions of Fishburn and of Keeney and Raiffa to justify a QALY-index

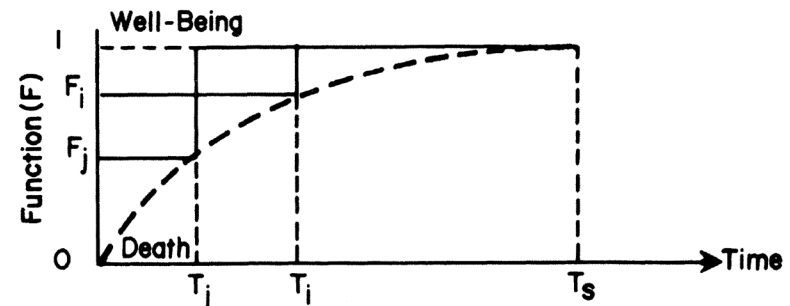
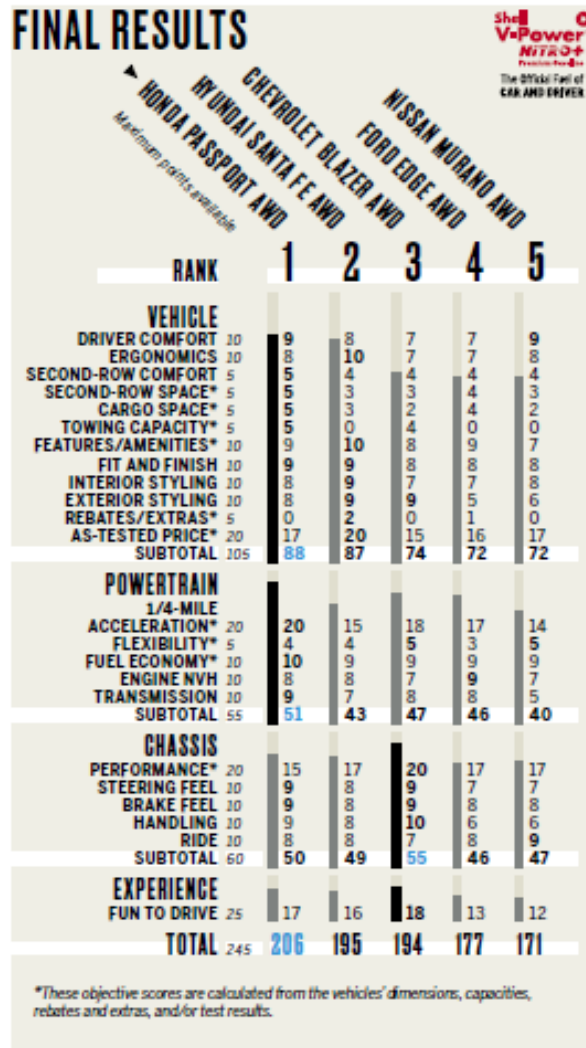


Fig. 4. Weighting the functional states through equivalence in time (group size is the same for all times).

Conversational applications



Comparison of vehicles from the magazine Car and Driver

- Notice the implied hierarchy of objectives: Vehicle, Powertrain, Chassis, Experience
- Notice the implied weights on the attributes: more points to the more important ones
- The summation provides the approximate MAU values for rankings

Observations from MAU theory:

- Objectives and attributes should meet appropriate independence conditions and should not be redundant
- Implied weights should reflect the ranges over which the attributes are measured

MAU provides a coherent intellectual basis for common sense applications

Research in Probability Assessment:

- Super High Church:

- DeFinetti (1937): If You are “coherent,” You have probabilities.

Coherence. It is assumed that You do not wish to lay down bets which will with *certainly* result in a loss for You.† A set of your previsions is therefore said to be *coherent* if among the combinations of bets which You have committed yourself to accepting there are none for which the gains are *all uniformly negative*.‡

From DeFinetti (1974)

- Savage (1954): Given certain axioms, you have probabilities and utilities.
 - Anscombe, Aumann (1963): Given other axioms, you have probabilities.
- Early DA Research:
 - Winkler (1967, JASA): “Despite the importance of prior distributions in Bayesian analysis, little previous work has been done on the practical problems of the assessment of non-diffuse distributions.”
 - Proposed and tested methods for assessing distributions
 - Winkler (1968, MS): “If a problem is important enough to warrant consulting an expert, it may be important enough to warrant consulting more than one expert.”
 - Proposed and tested methods for combining distributions

Best practices for assessment:

- Spetzler and Stael von Holstein (1975):

MANAGEMENT SCIENCE
Vol. 22, No. 3, November 1975
Printed in U.S.A.

Exceptional Papers

PROBABILITY ENCODING IN DECISION ANALYSIS††*

CARL S. SPETZLER AND CARL-AXEL S. STAËL VON HOLSTEIN§

Stanford Research Institute

This paper presents the present philosophy and practice used in probability encoding by the Decision Analysis Group at Stanford Research Institute. Probability encoding, the process of extracting and quantifying individual judgment about uncertain quantities, is one of the major functions required in the performance of decision analysis. The paper discusses the setting of the encoding process, including the use of sensitivity analyses to identify crucial state variables for which extensive encoding procedures are appropriate. The importance of balancing modeling and encoding techniques is emphasized and examples of biases and unconscious modes of judgment are reviewed. A variety of encoding methods are presented and their applicability is discussed. The authors recommend and describe a structured interview process that utilizes a trained interviewer and a number of techniques designed to reduce biases and aid in the quantification of judgment.

- Suggestions include:
 - 30-90 minute interview process; motivate the questions
 - Structure assessments carefully
 - Use reference gambles (would you rather bet on blue or the event?) rather than directly asking “what is your probability?”

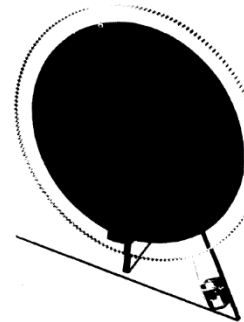


FIGURE 1. Probability Wheel.

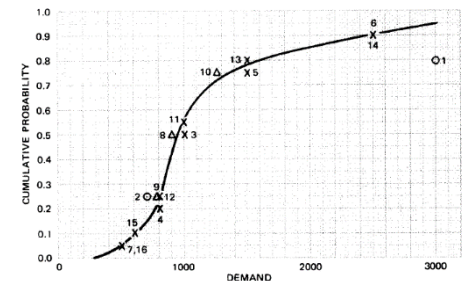


FIGURE 3. Example of a Curve Fitted to Responses.

Calibration work in MS:

- Harrison (1977) Independence and Calibration in Decision Analysis
- Wallsten, Budescu (1983) Encoding Subjective Probabilities:
A Psychological and Psychometric Review

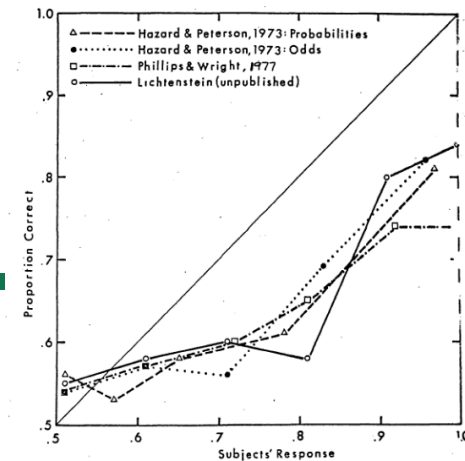
State of the Art

ENCODING SUBJECTIVE PROBABILITIES: A
PSYCHOLOGICAL AND PSYCHOMETRIC REVIEW*

THOMAS S. WALLSTEN† AND DAVID V. BUDESCU‡

Experts versus Nonexperts

The clearest difference to be seen between these two groups is in terms of the calibration studies. When encoding subjective probabilities about events with which they are familiar, experts can be exceedingly well calibrated, whereas a similar degree of goodness has rarely been demonstrated by nonexperts in laboratory contexts. Nonexperts do show relatively rapid, but limited improvement in calibration with training and feedback, and by the same token experts are less well calibrated when required to assess events defined differently from what they are used to considering.



Lichtenstein, Fischhoff, Phillips (1981) "Calibration of Probabilities: The State of the Art to 1980"

- Ravinder, Kleinmutz, Dyer (1988) The Reliability of Subjective Probs. Obtained Through Decomposition
 - Kahneman and Lovello (1993) Timid Choices and Bold Forecasts
 - Wallsten, Budescu, Zwick (1993) Comparing the Calibration & Coherence of Numerical & Verbal Prob.
 - Clemen, Fischer, Winkler (2000) Assessing Dependence: Some Experimental Results
 - Fox, Clemen (2005) Subjective Prob. Assessment in DA: Partition Dependence and Bias ...
 - Clemen, Ulu (2008) Interior Additivity and Subjective Probability Assessment of Continuous Variables
 - Walters et al. (2017) Known Unknowns: A Critical Determinant of Confidence and Calibration
 - Tannenbaum et al. (2017) Judgment Extremity and Accuracy Under Epistemic vs. Aleatory Uncertainty
 - Reigner (2018) Probability Forecasts Made at Multiple Lead Times
- + A lot of work on "scoring rules" for evaluating probabilities and on combining forecasts**

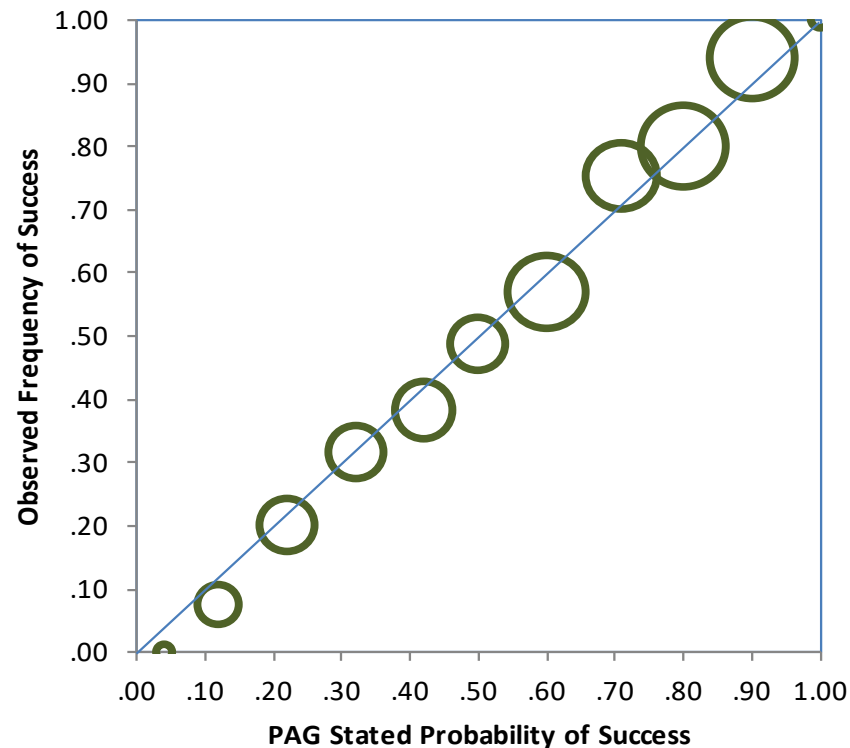
One thread in this work on calibration:

- Spetzler and Stael von Holstein (1975):
Begin by asking the subject for what he considers to be extreme values for the uncertain quantity. Then ask for scenarios that might lead to outcomes outside of those extremes. ...
- Kahneman and Lovello (1993): Consider an unnatural “outside view” that “avoids the details of the case at hand.”
- Tannenbaum et al. (2017): Compare
 - *The Chicago Bulls will play the Detroit Pistons on March 21st. What is the probability that the Bulls will win?*
 - *The Chicago Bulls will play the Detroit Pistons on February 20th, March 21st, and April 3rd. What is the probability that the Bulls will win on March 21st?*
- **Recurring Takeaway:** Encourage people to consider the full set of possibilities rather than a specific and compelling story.

Eli Lilly's Experience (data provided by Jay Andersen and Charles Persinger, Eli Lilly and Company)

- At Lilly, an independent board (12-15 members) has assessed the prob. of technical success for most R&D projects since 1997.
- One board (the PAG) is responsible for the whole R&D portfolio.
- Process led by a facilitator; board members have been trained and have access to historical results.
- Assessments for all stages of development:
 - $P(\text{preclinical success})$
 - $P(\text{phase 1 success given preclinical success})$
 - $P(\text{phase 2 success given phase 1 success})$
 - $P(\text{phase 3 \& registration success given phase 2 success})$
- Differing assessments are “averaged” by the chair of the PAG.
- In a retrospective study, Lilly compared probability estimates to outcomes for 1274 PAG estimates from 1997-2019.

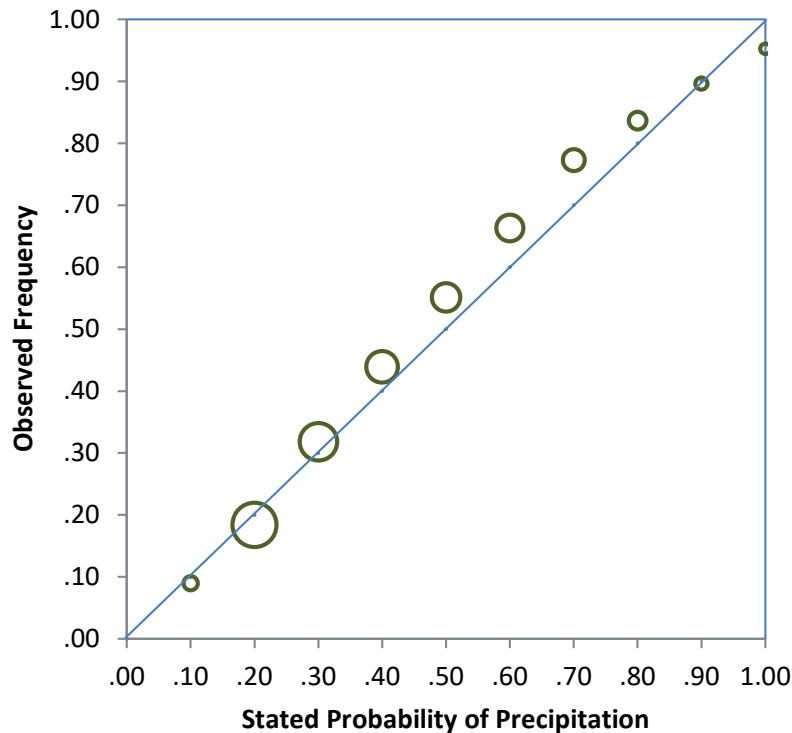
Eli Lilly results:



Conclusion: *“Our experience has shown that a well-planned process for probability assessment can provide executives with reliable measurements of technical feasibility ... Probability is an excellent language for quantifying this uncertainty.”*

Performance of NWS forecasters:

(data provided
by Eric Bickel)



A calibration plot for U.S. National Weather Service Forecasters for day ahead Probability of Precipitation Forecasts, for the "warm season" (April-September). From November 2008-October 2010. (Averaged over all regions in the US.)

There are 248,348 observations. Circle sizes are proportional to the frequency of the stated forecasts.

Note: Typically, no forecast is issued if the probabilities are low; thus there is no data for 0% and little for 10% probabilities.

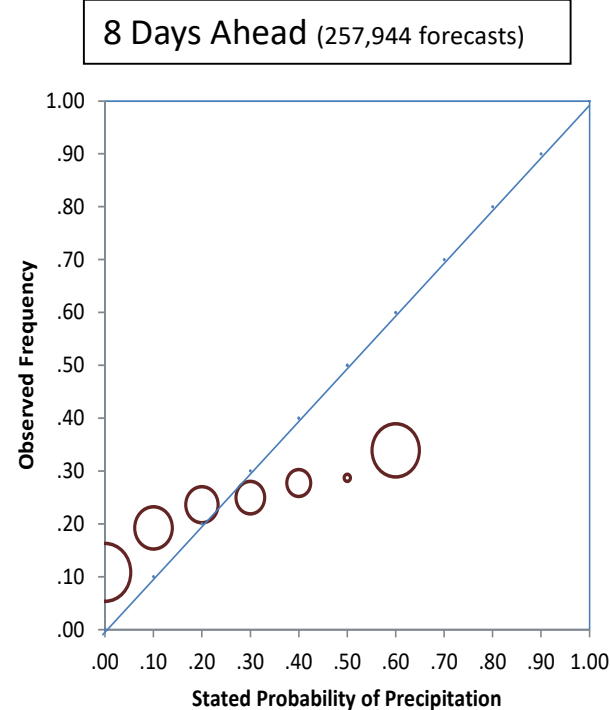
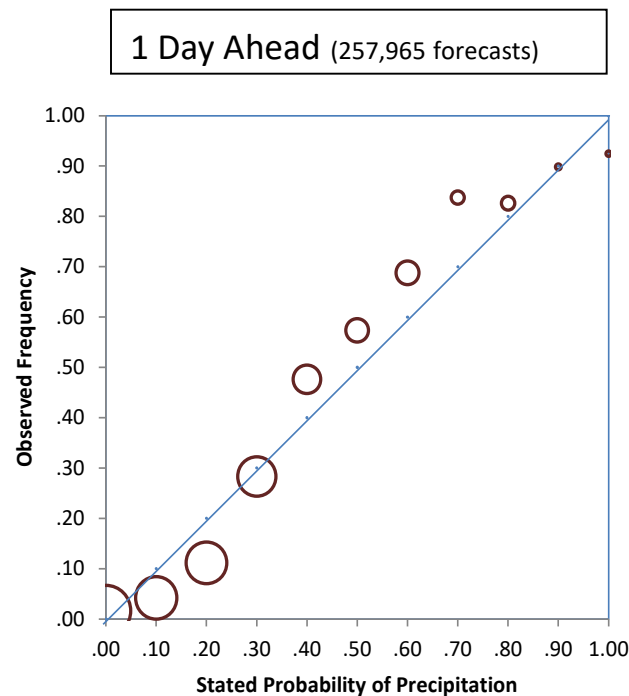
Source: Bickel, Floehr, Kim (2011)

Magazine

The Weatherman Is Not a Moron

By NATE SILVER SEPT. 7, 2012

Performance of The Weather Channel:



Same time frame and locations as NWS forecasters.

Idiosyncratic features:

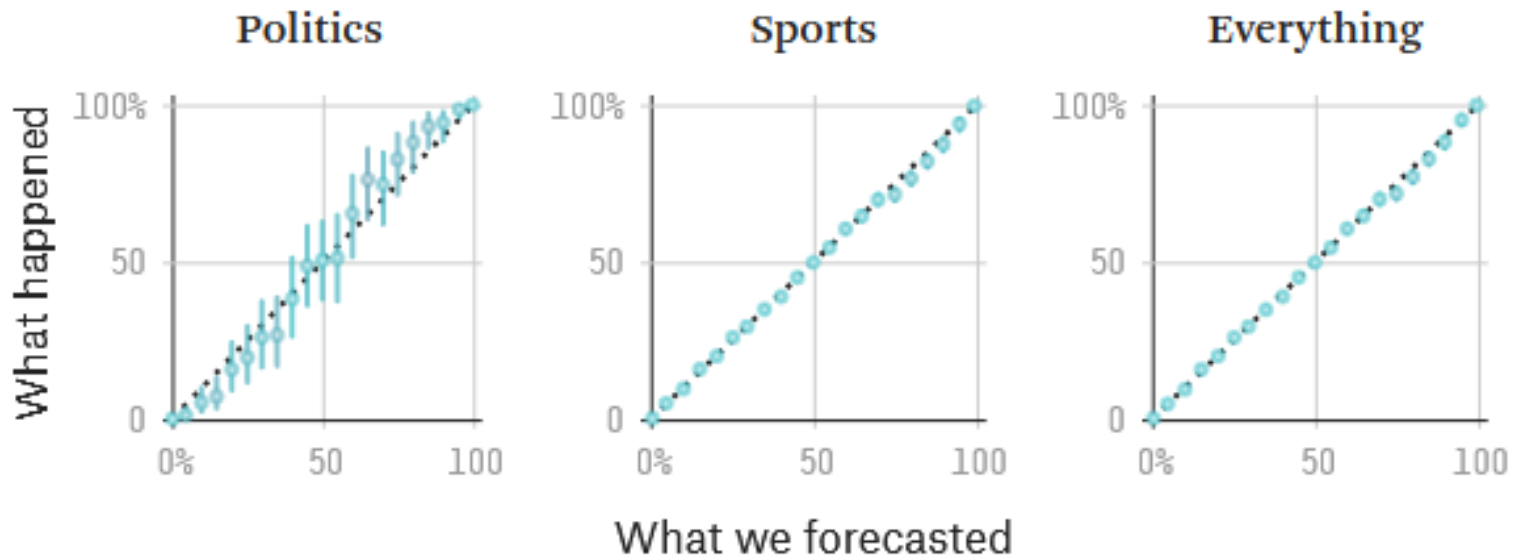
For near-term forecasts, low probabilities of precipitation (10-20%) are overstated.

Long-term forecasts are very poorly calibrated: note that they (almost) never say 50%!

National Weather Service forecasts do not have these issues.

Source: Bickel, Floehr, Kim (2011)

How Good Are FiveThirtyEight Forecasts?



FiveThirtyEight uses sophisticated Bayesian models to combine polling data in its political forecasts and uses other models for sports forecasting.



Conclusions:

- High-church research on DA theory and methodology – including that published in *MS* – has paid dividends in low-church and conversational applications.
 - DA thinking remains important in a “big data” world.
- Many low-church applications are outside of the scope of *MS* and may not be visible to the *MS* academic community.
- Conversational applications are of considerable worth:
 - People thinking clearly about tradeoffs
 - People talking clearly (using probabilities!) about uncertainty
 - Nudges!
- Purveyors of “conversational” advice should be aware of and sensitive to the concerns identified by high-church researchers (and *vice versa*).

What new? What's next?

- Behavioral decision research is “winning” at MS. Examples:
 - Budish, Kessler (2022) Can Market Participants Report Their Preferences Accurately?
 - He (2021) Revisiting Ellsberg's and Machina's paradoxes: ...
 - Baucells, Zhao (2020) Everything in Moderation: Foundations ... of the Satiation Model
 - Baillon, Bleichrodt, Spinu (2020) Searching for the Reference Point
 - Li, Muller, Wakker, Wang (2018) The Rich Domain of Ambiguity Explored
- We would like to see more DA-based recommender systems, combining (big) data with clear preferences and probabilities
 - Yelp, Pandora, Spotify, Netflix, ... with preference inputs?
 - Financial Engines for retirement planning
 - Budish, Kessler (et al.)'s “Course Match” system
 - School choice advisors?
 - COVID testing app with Bayesian interpretation?
 - Customized car-buying advice from, say, Car and Driver