# Book Review

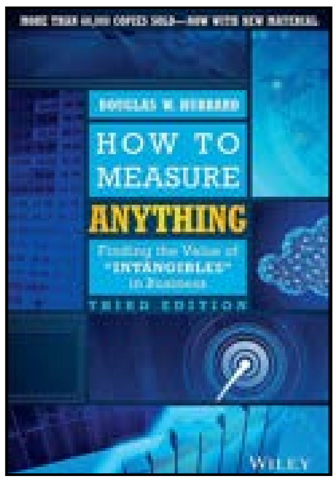
**Lawrence D. Stone**
Metron, Inc.
Reston, Virginia, USA
stone@metsci.com

## How to Measure Anything: Finding the Value of "Intangibles" in Business

**Douglas W. Hubbard**
3rd Edition, Wiley, 2014
ISBN: 978-1-118-53927-9, $52.00 USD, Hardcover: 432 Pages

## INTRODUCTION

The provocative title of this book caught my eye since I was embarking on a new project that involved measuring the operational utility of a system. How do you measure that? I hoped this book would help me answer that question. As one can see from the title, the book is directed toward business problems, but the methods Hubbard proposes can be applied to any problem where one is using subjective or expert opinion to estimate inputs to a problem.




The author does an excellent job of drawing the reader into his book which involves much more than just measurement. First, he explains what he means by "measure." He means make an estimate. He points out that all measurements are estimates, even ones made with highly accurate and precise instruments. They all have measurement errors or uncertainties. In fact, we know from the Heisenberg uncertainty principle that it is theoretically impossible to measure some quantities, such as position and momentum, with no error. Hubbard points out that the goal of measurement is to reduce the uncertainty in the quantity being measured, that is, to improve the estimate of the quantity. The author is serious about his claim of being able to measure anything. He gives some examples to back-up his claim. In one situation, the Cleveland Orchestra wanted to measure whether its performances were improving. How did they do this? By counting the number of standing ovations for its performances. If this number increased, the orchestra felt that their performances must be improving.

He then asks, "Why do you want to make the measurement?" Usually, the reason is to help make a decision in the presence of uncertainty. This can be a business decision, such as whether to upgrade an information technology system or launch a new product, but this can be true for almost any difficult decision. More generally, the author considers the situation where there is a decision to make, and there is a model that predicts the outcomes of the possible decisions. However, the outcomes and associated risks are uncertain, typically because some of the inputs to the model are uncertain. To make a good decision, one must reduce these uncertainties.

As the book progresses, the reader realizes that the author has led him from the provocative claim of being able to measure anything to a Bayesian decision theory approach to making decisions.

The book is intended for industry, government, or civic organizations that want to make better decisions. It is not aimed at the data fusion community which is already well-aware of Bayesian decision theoretic methods. Even so, I found the book had useful nuggets for the data fusion community. In terms of estimating "unknown" parameters, the author claims that the following are almost always true, even if the problem is "unique and unlike any other problem ever encountered," which is a claim often made by decision makers.

- It has been measured before.
- You have far more data than you think.
- You need far less data than you think.
- Useful, new observations are more accessible than you think.

The author gives examples for each of these claims. For the first claim, he suggests searching the internet for references to papers or documents that present results for measuring the item in question or perhaps something closely related. He provides suggestions for how to form queries to get specific rather general information. As researchers, most of us are familiar with this approach when it comes to writing a paper. We use a search engine and references in papers to find prior publications related to our work. Following the author's lead, I performed a search on "operational utility" and got measurements designed for utilities such as electric or gas ones. I then refined my search to "measuring military operational effectiveness" and obtained a much more useful set of references.

To support the claim that you need far less data than you think, he first observes that in most cases, many of the inputs have very little effect on the estimated results of a model, or they are known with enough certainty that they are not driving the risk in the decision. It is common that one or two very uncertain inputs are producing the uncertainty in the results. He then points out that a small number of measurements of these inputs can reduce their uncertainty dramatically. To support this claim, he provides the following rule of five: *There is a* 93.75% *chance that the median of a population is between the smallest and largest values* ***in any random sample of five*** [emphasis added] *from the population.*

Of course, there are some caveats, e.g., the samples must be random with each sample being an independent draw from the population. However, this is a remarkable result showing how

***"…managers are reluctant to measure inputs that are important to decisions, believing in many cases that it would be too expensive or too hard to make the measurements that would reduce the uncertainty in a decision."***

a small amount of data can produce a large amount of information (reduction in uncertainty) when your initial uncertainty is large. If the initial uncertainty is small, the reduction will be much smaller, but you may not need more information about an input with small uncertainty.

The author gives another striking example called the "Urn of Mystery." This example is an idealization of the situation where you wish to estimate the percentage of a population that has a certain characteristic, like the percentage of employees of a large company that take public transportation to work. The mystery is described as follows: *Suppose I have a warehouse full of large urns. Each urn is filled with marbles and each marble is either green or red. The percentage of marbles that are green in a single urn can be anything from 0 to 100% and all percentages are equally likely. The rest of the marbles are red. Assume the marbles are thoroughly and randomly mixed in each urn. Suppose I draw one urn at random, draw one marble from that urn, and observe its color. What is the probability of that color being the majority color over all the urns in the warehouse? The answer is 75%.* (This is correct, I checked the math myself.) Here we have obtained a remarkable amount of information about that population from only one sample!

These are interesting examples, but why are they important? The author has helped many companies make important decisions. In many of these cases, he observed that managers are reluctant to measure inputs that are important to decisions, believing in many cases that it would be too expensive or too hard to make the measurements that would reduce the uncertainty in a decision. The point of the examples is to show that when uncertainty is the largest, you obtain the most benefit from a few measurements. This reluctance to measure is not limited to business people. It also occurs among decision analysts. Hubbard wrote an article titled "Modeling without Measurements" for the October 2009 edition of *ORMS Today* [1]. He noted that: *A detailed analysis of 60 major decision analysis projects, a survey of Monte Carlo users, and a review of related literature showed…*

1. *Inputs for models are rarely calibrated. Models that depend heavily on subjective estimates [rarely employ] methods to adjust for errors [in these estimates].*
2. *Modelers rarely improve the initial model [by performing] empirical measurements of uncertain values….*
3. *Even organizations steeped in performance metrics rarely measure the performance of [the] models themselves.*

I don't know about you, but I winced when I read 1–3 above. It is too true, even of my own work. For example, some of my current projects rely on subjective estimates from experts. Being good Bayesians, my coworkers and I are developing methods for these experts to express their estimates and their uncertainty in those estimates. However, until I read Hubbard's book, I had not thought of how to calibrate and improve those estimates. Here he provides a real nugget for Bayesians and modelers in general.

Hubbard quotes Daniel Kahneman and Amos Tversky, who performed many studies testing people's probabilistic intuition, as saying: *Our thesis is that people have strong intuitions about random sampling; that these intuitions are wrong in fundamental respects; that these intuitions are shared by naïve subjects and by trained scientists; and that they are applied with unfortunate consequences in the course of scientific inquiry.*

For example, in [2], Kahneman and Tversky showed that people will routinely overestimate the probability of extreme sample results. These results emphasize the need to calibrate experts' probability judgements.

## CALIBRATING EXPERTS

Chapter 5 provides a method for calibrating the uncertainty estimates of experts. Hubbard begins this chapter with the following discussion: *How many hours per week do employees spend addressing customer complaints? How much would sales increase with a new advertising campaign? Even if you don't know the exact values to [answer] questions like these, you still know something. You know that some values would be impossible or at least highly unlikely. Knowing what you know now about something actually has an important and often surprising impact on how you should measure it or even whether you should measure it. In fact, quantifying our current level of uncertainty is a key part of the statistical methods we are using…*

The author's calibration method involves asking the experts to make estimates in terms of 90% containment intervals of items such as:

- In what year did Isaac Newton publish the universal laws of gravitation?
- How many inches long is a typical business card?

The experts are provided with the answers so they can compare their estimates to the actual values to see whether their intervals are typically too large or too small. This process and the subsequent feedback to the experts was repeated a number of times to allow the experts to improve their uncertainty estimates. Does this improve the experts' ability to express their uncertainty for an unrelated question such as one that is important for a business decision? Hubbard performed the following study to find out.

Hubbard reported the results from 927 subjects (see pp. 112 and 113 of the book) that he put through a half day of calibration tests of the type described above. The initial test contained 10 questions. On this initial test, only 50% of the 90% containment intervals contained the correct answer. At the end of the half day, he gave a final 20 question calibration test. For this test, over 80% of the intervals contained the correct answer. Clearly a big improvement. The tests were given to different subjects over several years, and the test questions varied from year to year. The results quoted above are aggregated over all the subjects that finished the half day of calibration. Impressive results, but they required a half day of calibration. If the decision is important, this effort should be worth it. The book's Appendix provides a set of calibration questions and answers that may be used to calibrate experts.

*"Often only a few measurements will produce a dramatic reduction in uncertainty and thus risk."*

## MEASURING THE PERFORMANCE OF MODELS

In my experience, this does happen occasionally but not as often as it should. My work has been predominately with the US Navy, so my examples are drawn from that experience. The Navy does test some detection and tracking systems in the following way. Developers are given some real (at sea) data to test and develop their system. This provides a way for the developers to adjust the inputs to their systems to improve their performance. The software implementing the systems is given to a third party to test and grade their performance on a hidden data set. If the highest performing system performs significantly better than the existing system, the existing system is replaced by the new one.

This is not quite the same as a test during a real operation, but such tests are often hard (if not impossible) to perform. I have also developed a Bayesian search and rescue planning system for the US Coast Guard. The system is called SAROPS which is presently in use by the Coast Guard and is highly regarded by them. However, we have never performed a test of SAROPS vs the Coast Guard's manual planning method. It is hard to imagine how you would do this. You could contemplate flipping a fair coin before a search to decide whether the search should be planned with SAROPS or the manual method and then compare results. Of course, for ethical reasons, you would never perform this experiment. However, Hubbard cites a number of studies [3],[4], showing that even rough statistical models greatly outperform seat of the pants decision making. That is encouraging.

I do know of one case where it was possible to compare a Bayesian search planning methodology to manual planning by experts. In the 1970s, the Navy routinely flew aircraft to drop sonobuoys to detect adversarial submarines whose rough location was provided by a surveillance system. My colleagues and I at Daniel H. Wagner Associates developed a Bayesian search planning program for use in planning where to drop the buoys. Once the system was developed, it was used experimentally on some searches while others were planned using the existing manual methods. At the end of this test period, we compared the results from the two methods. The Bayesian computer method increased detection probability by a factor of 2 even though the computer method was used on the harder, more complicated problems, while the manual method tended to be used on the simpler, easier problems. When people ask me whether Bayesian search planning actually works (improves the results), I usually give them this example. However, test results like these are rare for military systems for obvious reasons.


### *PART I. THE MEASUREMENT SOLUTION EXISTS*

1. The Challenge of Intangibles
   Intangibles are measurable
2. An Intuitive Measurement Habit
   How to estimate and perform simple experiments
3. The Illusion of Intangibles: Why immeasurables aren't
   Concept, object, and methods of measurement

### *PART II. BEFORE YOU MEASURE*

4. Clarifying the Measurement Problem
   What uncertainty and risk mean
5. Calibrated Estimates: How much do you know now?
   How to calibrate experts and why calibration is useful
6. Quantifying Risk through Modeling
   Monte Carlo methods of quantifying risk
7. Quantifying the Value of Information
   The value of uncertainty reduction

### *PART III. MEASUREMENT METHODS*

8. The transition from what to measure to how to measure
9. Sampling reality: How observing some things tells us about all things
10. Bayes: Adding to what you know now

### *PART IV. BEYOND THE BASICS*

11. Preference and Attitudes: The softer side of measurement
12. The Ultimate Measurement Instrument: Human judges
13. New Measurement Instruments for Management
14. A Universal Measurement Method: Applied information economics

***Appendix. Calibration Tests (and their Answers)***

## LIST OF CHAPTERS

From the list of chapters (text box), one can obtain an outline of the author's method of convincing business managers to use statistical decision tools when making hard decisions. As you can see, the theme of the book is measurement and how to measure almost anything. However, in Part III, he introduces the concept of decision models by using them to express the risk (and reward) of making a decision. The topic of risk leads naturally to the concept of reducing uncertainty to reduce risk and to allow a manager to make a better decision. One of the best ways to reduce uncertainty is to make some measurements. As the author notes, often only a few measurements will produce a dramatic reduction in uncertainty and thus risk.

This book is long, roughly 400 pages, because the author provides lots of discussion, anecdotes, and examples from his long experience helping managers make difficult and important decisions. This is the right approach for managers. He builds the case for statistical decision making slowly using many examples and lots of discussion. He leads the manager slowly from wanting to measure something to statistical decision analysis. In the end, this is a full-blown discussion of statistical decision analysis for managers with practical instructions for performing and using these analyses. Quite a performance. All with a minimum of mathematics. Very impressive and an interesting read.

## NUGGETS

I gleaned three nuggets from this book which I think are useful for data fusion analysts as well as decision analysts:

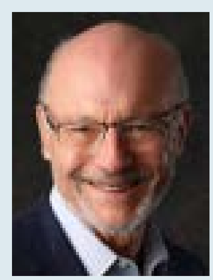
1. If you are using subjective inputs provided by experts, calibrate them. This may be difficult or time consuming to do, but if the decision or the system relying on these inputs is important, it is probably worth doing this or at least trying to do it.
2. If some inputs or parameters are driving the uncertainty and, therefore, the risk in a model, try to devise some method of obtaining measurements of these inputs to reduce their uncertainty. As the author points out, if the uncertainty in a value is large then a few measurements can greatly reduce it.
3. Try to test the effectiveness of your model or decision system. If you can't do this in an operational situation, perhaps you can use recorded data to test your system. Of course, withhold some data to perform a blind test of the system after you have tuned it on the rest of the data.

As well as being an interesting book to read, I learned some interesting facts about small samples and found several items of good advice for my own work. Now let's see if I can follow that advice.

## REFERENCES


1. Hubbard, D. How the Decision Analysis Culture's Lack of Empiricism Reduces its Effectiveness. *Modeling Without Measurements*. (2009). https://doi.org/10.1287/orms.2009.05.12
2. Kahneman, D., and Tversky, A. Subjective Probability: A Judgement of Representativeness. *Cognitive Psychology*, Vol. 3, (1972), 430–454.
3. Grove, W.M. and Meehl, P.E. Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical Statistical Controversy. *Psychology, Public Policy, and Law*, Vol. 2, (1996), 293–323.
4. Grove, W.M., Zald, D.H., Lebow, B. S., Snitz, B. E., and Nelson, C. Clinical versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, Vol. 12, No. 1, (2000), 199–30.





**Lawrence Stone** is Chief Scientist at Metron Inc. He is a member of the National Academy of Engineering and a fellow of the Institute for Operations Research and Management Science (INFORMS). He is a recipient of the Jacinto Steinhardt Award from the Military Applications Section of INFORMS in recognition of his applications of Operations Research to military problems.

In 1975, the Operations Research Society of America awarded the Lanchester Prize to his text *Theory of Optimal Search*. In 1986, he produced the probability maps used to locate the *S.S. Central America* which sank in 1857, taking millions of dollars of gold coins and bars to the ocean bottom one-and-one-half miles below. In 2010, he led the team that produced the probability distribution that guided the French to the location of the underwater wreckage of Air France Flight AF447. Recently, he used search theory methods to help guide the Canadian exploration company, Aurania, to the location of one of the lost Spanish gold cities in Ecuador.

He coauthored the 2016 book *Optimal Search for Moving Targets*. He was one of the primary developers of the Search and Rescue Optimal Planning System (SAROPS) used by the U.S. Coast Guard since 2007 to plan searches for people missing at sea. He continues to work on a number of detection and tracking systems for the U.S. Navy. He is a coauthor of the 2014 book *Bayesian Multiple Target Tracking, Second Edition* as well as the 2023 book *Introduction to Bayesian Tracking and Particle Filters*.