LA-UR-00-1766

Approved for public release; distribution is unlimited.

Title: Testing the Untestable: Reliability in the 21st Century

Author(s): Thomas R. Bement, TSA-1
Jane M. Booker, TSA-1
Sallie Keller-McNulty, TSA-1
Nozer Singpurwalla, The George Washington University

Submitted to: AAAS Science

Los Alamos

NATIONAL LABORATORY

Testing the Untestable: Reliability in the 21st Century

Thomas R. Bement,[1] Jane M. Booker,[1]
Sallie Keller-McNulty,[1] and Nozer D. Singpurwalla[2]

As science and technology become increasingly sophisticated, government and industry are relying more and more on science's advanced methods to determine reliability. Unfortunately, political, economic, time, and other constraints imposed by the real world inhibit the ability of researchers to calculate reliability efficiently and accurately. Because of such constraints, reliability must undergo an evolutionary change. The first step in this evolution is to reinterpret the concept so that it meets the new century's needs. The next step is to quantify reliability using both empirical methods and auxiliary data sources, such as such as expert knowledge, corporate memory, and mathematical modeling and simulation.

[1]Los Alamos National Laboratory
[2]The George Washington University

1.0 Introduction

"Reliability" is a charged word guaranteed to get attention at its mere mention. Bringing with it a host of connotations, reliability, and in particular its appraisal, faces a critical dilemma at the dawn of a new century.

Traditional reliability assessment consists of various real-world assessments driven by the scientific method. In other words, conducting extensive real-world tests over extensive time periods (often years) enabled scientists to determine a product's reliability under a host of specific conditions.

In the 21st century, humanity's technological advances walk hand in hand with myriad testing constraints, such as political and societal mores, economic and time considerations, and lack of scientific and technological knowledge. Because of these constraints, the accuracy and efficiency of traditional reliability becomes much more questionable.

For example, how can traditional reliability assessment techniques determine the dependability of manned space vehicles designed to explore Mars, given that humanity has yet to venture that far into space? How can we determine the reliability of a nuclear weapon, given that the world has in place test-ban treaties and international agreements? And finally, how can we decide which artificial heart to place into a patient, given neither has ever been inside a human before?

To resolve this dilemma, reliability must be (1) reinterpreted and (2) quantified. To reinterpret reliability, we must first move away from logical inferences and move closer to empirical evidence. The primary reason for this shift is because logic encompasses a world of tautologies, with terms such as "certainty" and "impossibility." Techniques driven by logic calculate numbers such as 10^{-9} for failure rates. Does this number mean one failure in 10^{9} identical trials? Is it possible in the real world to create identical trials? From a practical point of view, logical approaches are much too abstract to be an effective means of determining the reliability of products in the real world.

As has been previously stated, empirical evidence drives the traditional meaning of reliability. Using the scientific method, researchers use empirical evidence to determine the probability of success or failure. Therefore, reliability can be seen as a mirror image of probability. But what exactly is probability, particularly at the dawn of a new century? The first part of this paper presents an overview of several interpretations of probability and how they relate to reliability.

Once reliability has been reinterpreted, we must next quantify it. And this is where advanced methodologies mix with traditional ones. Rather than relying alone on so-called "hard data," the redefined concept of reliability incorporates auxiliary sources of data, such as expert knowledge, corporate memory, and mathematical modeling and simulation. By ~~"fusing"~~combining both types of data, reliability assessment is ready to enter the 21st century.

2.0 Reliability, Probability, and Decision Making

In this section we discuss reliability's link to decision making and its close association with probability. This overview will serve as a foundation for the next section, which addresses the reinterpretation of reliability.

2.1 What is Reliability?

When most individuals think of the term reliability, they equate feelings of credibility, trustworthiness, and dependability. Some specialists (e.g., social scientists) have a much narrower interpretation, one in which reliability equates with the consistency of a test instrument (such as psychological test). A classical example of this interpretation is the rising and setting of the sun. Because humanity has seen the sun rise and set for as long as it can remember, there is an almost certain belief that the sun will rise and set tomorrow.

This paper defines reliability as a mathematical term (see Barlow and Proschan, 1975). Thus, reliability is a quantified measure of uncertainty about a particular type of event (or events). ~~Thus, reliability~~Reliability can be seen as a ~~type~~ function of probability. In the sun example~~ given above~~, it therefore is highly probable that the sun will rise and set tomorrow, given the wealth of empirical data.

2.2 Reliability's Role in Decision Making

Reliability can be seen as a tool that helps individuals make logically sound decisions. Decisions can be technical (e.g., science and engineering) or nontechnical (e.g., strategy or management). This role brings up two principal questions:

- If reliability is defined as a quantified measure of uncertainty, then whose uncertainty is it?
- What does it mean to say that a decision is logically sound?

There are several answers to the first question. For example, the uncertainty may pertain to a particular group of individuals or there may be an inherent notion of "universal" uncertainty. Section 45.0 shows that the answer to this first question dictates the paradigm used to quantify reliability.

Logically sound decisions use what is called a normative approach. This approach involves a system of rules (also called axioms) that a decision maker or a group of decision makers has agreed upon as being appropriate. This type of approach in essence tells us how we *should* act, not how we *actually* act. In most instances, individuals make decisions based on emotion, whim, and personal/political agendas. A classic example of such decision making is when individuals elect to drive instead of taking an airplane, despite empirical evidence that the latter is a much safer mode of transportation.

However, the normative approach is crucial in decision making because of its logic, which in turn makes such decisions much easier to explain and, if necessary, justify.

Figure 1 provides an outline of the normative approach in the form of what is known as a decision tree (this tree is a generic example for a system deployment decision).

A decision tree consists of one more decision nodes (shown as rectangles on Fig. 1) and one or more random nodes (shown as circles). A decision node always precedes a random node, but a random node may or may not be followed by a decision node. At the terminus of a tree ~~there~~ are ~~appear what are known as~~ "consequences~~.~~" ~~A consequence~~that can include

- tangibles such as costs, penalties and profits, and
- intangibles such as goodwill, tastes, and preferences.

When the consequences are quantified (and represented on a scale, such as from zero to one) they are known as "utilities."

The decision tree shown in Fig. 1 consists of one decision node from which spawn two decisions. If the system is deployed, then there are two possible outcomes, each of which brings about separate consequences. Similarly, when the system is not deployed, there is a resulting consequence. Although decision trees are not a new idea (for an overview of related literature, see Booker and Bryson, 1985a and 1985b), they remain a powerful characterization of the normative approach to decision making.

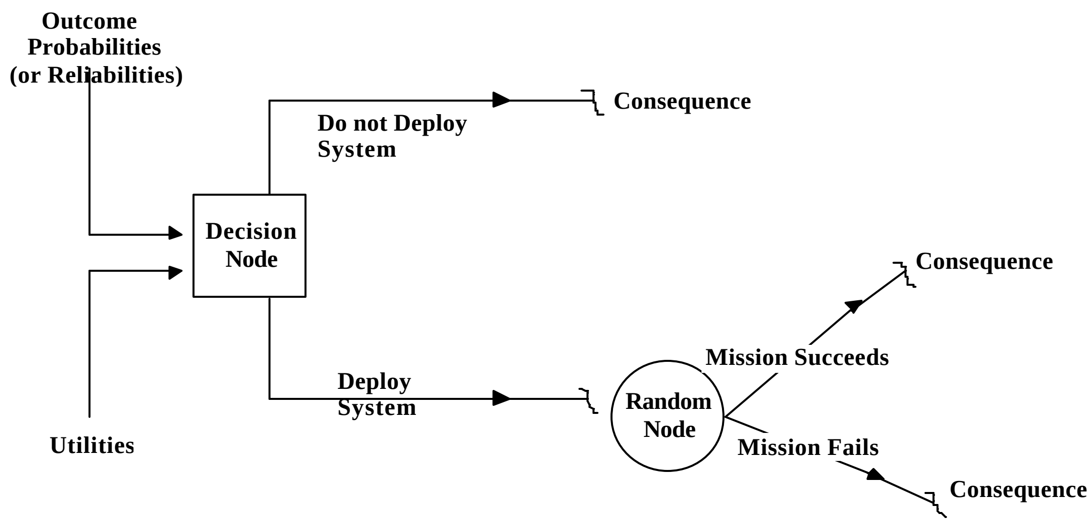

Fig. 1. An example of a decision tree.

There are two formal processes used to make decisions: the "Analytical Hierarchy Process" of Saaty (1980) and the "Statistical Decision Theory" (DeGroot, 1992).

Of these, only the latter is considered a normative approach. This theory in turn has two variants, the Bayesian and the frequentist, with the former considered normative. Commonly referred to as the Bayesian Decision Theory, it is entirely based on the calculus of probability (see Section ~~5~~3.0).

According to the dictates of this theory, a decision maker should choose that action which maximizes the expected utility~~—known as the "MEU principle"~~ [cf. Lindley 1985, p.59]. An action's expected utility is the sum of the products of the probability of an action's outcome and the consequence that results from the action-outcome combination. To compute the expected utility of each action, we must determine the probabilities of all the outcomes that can result from the action. Therefore, like utilities, outcome probabilities are the required ingredients of normative decision making in a certain class of problems, namely problems that involve the ability of systems to perform (or not) as desired.

2.3 Understanding Reliability

If we accept that reliability consists of a quantified measure of uncertainty, then we must once again ask to whom the uncertainty belongs: Does it belong to an individual? Is it an interpretation from a group? Or is it a notion of inherent universal uncertainty? This question's answer depends on the basis for the quantification of uncertainty. The basis could be based upon any of the following:

1. engineering, scientific, or other subject matter information;
2. mathematical models, physical models, and simulations;
3. informed testimonies and collective judgements from subject matter specialists;
4. corporate memory, commercial databases, knowledge bases, and historical information;
5. hard statistical data on several replicas of the uncertain event(s) of interest via experiments and tests; and

6. all the above.

This list represents numerous schools of thought. For example, some schools maintain that only hard statistical data are relevant for quantifying uncertainty, whereas other schools advocate that all the methodologies should be represented. Philosophical disagreement plays a critical role in quantifying reliability. For example, an individual who advocates hard data exclusively will calculate a reliability that is dramatically different from someone who advocates using hard data and modeling and simulation. The inherent universal uncertainty ~~is the aleatory uncertainty referred to earlier and~~ incorporates the random physical and natural variation left after exhausting the sources on information, data, and knowledge listed above. Our reinterpretation of reliability advocates the use of all data on the list, and its success has been demonstrated[1] (Meyer, Booker, and Bement, 1999).

3.0 Probability: A Method to Quantify Uncertainty (Reliability)

This section provides an overview of several attempts at addressing the quantification of uncertainty. These attempts have contributors with diverse backgrounds, from philosophers and economists to physicists, psychologists, sociologists, engineers, statisticians, and mathematicians. Each approach has it own merits and flaws. Before we discuss these approaches, we must introduce some notation and terminology.

3.1 Notation and Terminology

[1] PREDICT (which stands for Performance & Reliability Evaluation with Diverse Information Combination and Tracking) is an integrated reliability methodology that combines all available information, with appropriate uncertainties attached, relating to the system's performance. Information sources include expert judgement, historical data/information about the system's parts and processes, vendor/designer specifications, computer simulation output results, physical models, test data, and data on similar parts,

Let E, E_1, . . ., E_i, . . ., denote several uncertain events of interest at some reference time, say τ. Although it is common to set τ=o, it is important not to lose track of its presence. For example, E could denote an event that a deployed system accomplishes its mission. The complement of E is denoted by $\bar{E}$, the event that a deployed system fails to accomplish its mission (see Fig. 1). Another example could be that $E_i=\{T_i \geq t\}$, where T_i denotes the lifetime of the *i*-th sub-system of a deployed system (measured from the time of the system's deployment) and $\{T_i \geq t\}$ denotes the event that the *i*-th sub-system functions for at least t units of time. In this case, t is called the "mission time."

Let *H* denote the "history" or the "background information" available to the individual(s) contemplating the uncertain events, at time τ. In principle, *H* should encompass all that is known at time τ: scientific knowledge, engineering information, informed testimonies, design specification, physical models, computer codes, judgement, preferences, and hard historical data on replicates of the uncertain event (if available). Thus at any time τ, there is the known *H*, and the unknown E, E_1, . . ., E_i, . . ., .

The fundamental problem of the treatment of uncertainty is how the uncertainty about E, E_1, . . ., E_i, . . ., at the τ, should be quantified in the light of *H*. To address this problem, several approaches have been proposed, some of which pay attention to the issue of "whose uncertainty" and others which impose restrictions on what *H* can and cannot contain. Some of these proposed approaches are as follows:

- probability,
- belief functions,

---

subsystem, processes and systems. As new information becomes available, the method permits updating the performance of the system.

- possibility theory and fuzzy logic,
- upper and lower probabilities,
- Jeffrey's Rule of Combination,
- confidence limits,
- hypothesis testing with Type I and Type II errors,
- significance levels,
- maximum likelihood estimates, and
- goodness of fit tests.

Some of these approaches have a normative foundation, whereas others are *ad hoc*. We will focus on probability and make a case for it.

3.2 Probability and the Calculus of Probability

The calculus of probability consists of certain rules (or axioms) denoted by a number determined by $P^{\tau}(E; H)$, in which the probability of an event, E, is related to H at time τ. When the event E pertains to the ability to perform a certain function (e.g., survive a specified mission time), then $P^{\tau}(E; H)$ is known as the product's reliability. Therefore, reliability is de facto the probability of a certain type of an event.

When the item in question is a human subject, the term "survival analysis," rather than reliability, is commonly used. As indicated above, the mission time need not be measured in units of time, but rather it can be taken from other performance metrics, such as miles traveled, rounds fired, cycles completed, or output produced.

The calculus of probability consists of the following three rules: convexity, addition, and multiplication. These rules are given mathematically in order:

- $0 \leq P^{\tau}(E; H) \leq 1$, for any event E;
- $P^{\tau}(E_1, \text{ or } E_2; H) = P^{\tau}(E_1; H) + P^{\tau}(E_2; H)$ for any two events E_1 and E_2 that are mutually exclusive—that is, they cannot simultaneously take place; and

- $P^{\tau}(E_1 \text{ and } E_2; H) = P^{\tau}(E_1 \mid E_2; H) \cdot P^{\tau}(E_2; H)$, where $P^{\tau}(E_1 \mid E_2; H)$ is a quantification via probability of the uncertainty about an event E (supposing that event E_2 has taken place).

The quantity $P^{\tau}(E_1 \mid E_2; H)$ is known as the "conditional probability" of E_1, given n E_2. Note that conditional probabilities are in the subjunctive. In other words, the disposition of E_2 at time τ, were it to be known, would become a part of the history H at time τ. The vertical line between E_1 and E_2 represents a supposition or assumption about the occurrence of E_2. Finally, $P^{\tau}(E_1 \text{ and } E_2; H)$ also can be written as $P^{\tau}(E_2 \mid E_1; H) \cdot P^{\tau}(E_1; H)$ because at time τ both E_1 and E_2 are uncertain events and one can contemplate the uncertainty about E_1 supposing that E_2 were to be true or vice versa.

The calculus of probability does not interpret probability. ~~It does not do any of the following: it~~It neither tells use what probability means, nor is it concerned with issues, such as the nature of uncertainty, whose uncertainty, how large should H be, and how to determine $P^{\tau}(E; H)$ and make the result operational.

What does the calculus do? It simply provides a set of rules by which the uncertainties about two or more vents combine or "cohere." Any set of rules for combining uncertainties that are in violation of the rules given above are said to be "incoherent" with respect to the calculus of probability. The next section discusses why these rules are necessary.

3.3 Why Subscribe to the Calculus of Probability?

In addition to the calculus of probability, there are a number of methods designed to comb and specify uncertainties, such as

- Jeffrey's rule of combination (Jeffrey, 1983),

- possibility theory and fuzzy logic (Zadeh, 1979),
- upper and lower probabilities (Smith, 1961), and
- belief functions (Dempster, 1968).

With so many theories, why should anyone subscribe to the calculus of probability? In a book by Howson and Urbach (1989), a number of contributors attempt to answer this very question; contributors range from gamblers and philosophers to mathematicians, decision theorists, behavioral scientists, and experts in artificial intelligence and knowledge acquisition. In the following subsections, we present some of the arguments used to justify the calculus of probability.

3.3.1 The Flaw of "Can't Win"

Since the time of Cardono (who lived during the 1500s), gamblers recognized that avoiding the rules of probability in games of chance resulted in a "sure loss" (i.e., a Dutch-Book). A classic example is a coin toss in which the scenario is as follows: heads you lose, tails I win.

3.3.2 The Scoring Rule

To justify the calculus of probability, de Finetti (cf. Lindley, 1982) used a "scoring rule," which is used to ask an individual assessing and uncertainty to declare a number that best describes said individual's uncertainty. Once the uncertainty reveals itself or is resolved, the individual is scored (i.e., rewarded or penalized) according to how close the declared number was to reality.

De Finetti's core argument is that under some very general conditions, an individual faced with a collection of uncertainties must use the calculus of probability to maximize an overall score. The above claim is true for a large class of scoring schemes.

3.3.3 Betting Coefficients

In horse racing, certain numbers known as "betting coefficients" are used. These numbers are odds on or against a particular horse or horses involved in a horse race. It has been demonstrated (for example, by Howson and Urbach, 1989) that to maximize winnings (i.e., determine the most accurate probability of success), the betting coefficients must follow the calculus of probability.

3.3.4 Behavioristic Axioms

The three previous subsections provided answers based on gambling and scoring scenarios; such scenarios could be objectionable to some individuals, particularly those who question the moral and ethical implication of gambling and fierce competition. Therefore, Ramsey (1931) and Savage (1972) proposed an alternative system of "behavioristic axioms" to justify the calculus of probability. Based in mathematics, the Ramsey-Savage argument is related to coherence and consistency (an excellent exposition of this argument is given by DeGroot, 1970).

This argument has two principal difficulties. The first is that the intuitive and natural elements of gambling and scoring are lost; axiomatic arguments tend to be abstract and therefore less appealing. The second and perhaps more serious difficulty is that behavioristic axioms are in actuality violated by (most) individuals (cf. Kahneman, et al., 1986). Despite these criticisms, behavioristic axioms prescribe normative behavior. It is not imperative that the calculus of probability be treated as axiomatic, but rather that it is seen as a consequence of certain behavioristic axioms, with the latter being taken as given (not withstanding some criticism).

~~Our reinterpretation of reliability advocates the use of all data on the list. Based on this philosophical principle, we developed new methodology known as PREDICT[2] (Meyer, Booker, and Bement, 1999). PREDICT is an excellent example of a planning, tracking, and forecasting tool that reflects the needs of this fledgling century.~~

~~3~~4.0 Reinterpreting Reliability

~~According to Good (1965), there are approximately eleven ways to interpret probability.~~ As we have shown in Section 2.0, reliability is a probability, and because the latter can be interpreted in several ways, then it follows the reliability can be interpreted and quantified in several ways. Because of the flexibility of these terms, there are many different philosophies behind effective decision making. Thus, which philosophy we advocate will dictate the effectiveness of reliability assessment in the 21st century.

There are four principal theories related to interpreting probability:

- Classical Theory,
- A Priori or Logical Theory,
- Relative Frequency Theory, and
- Personalistic or Subjective Theory.

Although the interpretation of reliability has no effect on the calculus of probability (see Section ~~4~~3.0), the assignment of initial probabilities (which are needed to initiate the calculus) depends on reliability's interpretation. The following provide an overview of the key features of

~~[2] PREDICT (which stands for Performance & Reliability Evaluation with Diverse Information Combination and Tracking) is an integrated reliability methodology that combines all available information, with appropriate uncertainties attached, relating to the system's performance. Information sources include expert judgement, historical data/information about the system's parts and processes, vendor/designer specifications, computer simulation output results, physical models, test data, and data on similar parts,~~

the four theories mentioned above. For more detailed information, please consult Fine (1973), Good (1965), Maistrov (1974), Hacking (1974), or Gigerenzer et al. (1989).

~~subsystem, processes and systems. As new information becomes available, the method permits updating the performance of the system.~~

~~3~~4.1 Classical Theory

Influenced by Newton, the following "determinists" founded this theory: Cardano, Pascal, Fermat, Huygens, Bernoulli, DeMoiure, Bayes, Laplace, and Poisson.

As determinists, these individuals believed that every event, act, and decision was the inevitable consequence of antecedents that are independent of the human will. Of these individuals, the only one to venture a formal definition of probability was Laplace, who in essence described it as the ratio of favorable cases to the number of "equipossible" cases. Cases are equipossible if we have no reason to expect the occurrence of one over the other. The setup involving equipossible cases consists of the three following labels:

- "principle of indifference,"
- "principle of insufficient reason," or
- "Bayes' postulate."

Although this theory has merit in games of chance, it also has a number of ~~flaws~~deficiencies. For example, the principle of indifference appears to be a circular argument because equipossible implies "equiprobable" and vice versa. It also is difficult to divide up alternatives. For example, take the following problem:

> When rolling a die, what is the probability that it will land on "5"? The answer is 1/6, if the alternatives are considered 1, 2, . . . 6, but it is 1/2 if the alternatives are considered are a 5 and not a 5.

Perhaps the most crucial flaw is the potential for exceptions. For example, what if the die in the example above is loaded? What if there is a flaw in the die itself, which in turn affects the overall probability? Given these exceptions, now think of a unique space vehicle. What flaws will affect its reliability?

Although ~~flawed~~deficient, this theory is used to this day. It is particularly useful in teaching the concept of probability, as well as in a number of applications, such as Monte Carlo simulation, sampling, and experimental design.

~~3~~4.2 A Priori Theory

Although it was the economist Keynes who first proposed the A Priori Theory of probability, it was Carnap (a physicist whose interests ranged from logic and syntax to semantics and languages) who expanded upon it. Others who have contributed to this theory include Jeffreys (1961), Koopman, Kemney, and Good (1965), and Ramsey (1964).

The A Priori Theory is difficult to summarize in words because it involves the notions of logic and syntax. Basically, it interprets probability as a logically derived "entity." In other words, a violation of logic yields an inappropriate conclusion. Because it is difficult to apply this theory to reliability assessments, this theory is often discussed but rarely used.

~~3~~4.3 Relative Frequency Theory

Although the origin of this theory dates back to Aristotle, it was Venn who was the first to articulate the concept in 1866. Its mathematical development has been traced to von Mises (1957), whereas its philosophical discourse was developed by Reichenbach (1949). The following bullets summarize this theory's key elements:

- Probability is a measure of an empirical, objective, and physical fact of the real world. It is independent of human attitudes, opinions, models, and simulations. Von Mises believed probability to be a part of a descriptive model, whereas Reichenbach viewed it as part of the theoretical structure of physics.
- Because probability is based only upon observations, it can be known only *a posteriori* (literally, after observation).

4~~3~~.3.1 This Theory's Virtues

This theory applies to cases in which the indifference principle fails to hold (a six-faced die is loaded). Because the theory emphasizes observation, it has a strong link with the scientific method. Moreover, the theory rejects intangible things and relies on what many consider the essential tools of science: experimentation, observation, and confirmation through experimental replication.

4~~3~~.3.2 This Theory's Criticisms

The "core" of this theory is on replication. To achieve replication, we must

- introduce a random "collective" (i.e., a scenario involving events that repeat again and again),
- define that probability is indeed a random collective, and
- specify that probability is a property of the collective and not an individual member of said collective.

Creating a collective in the real world is a difficult problem. For example, tossing a coin an infinite number of times raises the following question: To be considered a collective, how similar must the tosses be? If the tosses are identical, then the outcome will not change. If they are dissimilar, how much dissimilarity is allowed (if this can be assessed at all)? Finally, relative frequency probability is never known, can never be known to exist (limits of sequences is an abstract mathematical notion), and its value can never be confirmed or disputed.

Although collectives can be developed for social phenomena (actuarial tables and individual IQs) and some topics in physics (e.g., movement of gas particles), it is for the most part a difficult task. Although the collective concept was first embraced by physicists like von Mises, it was subsequently rejected by individuals like Bohr and Schrodinger, both of whom were influenced by Heisenberg's "principle of uncertainty." This principal defined uncertainty and probability without the collective concept.

Under the relative frequency view of probability, τ and H have no role to play, so that $P^{\tau}(E; H)$~~$P^{\tau}(E; H)$~~ = P(E). Similarly, expert testimonies, corporate memory, mathematical models and scientific information do not matter; only hard data on actual events can be used to assess the initial probabilities.

4~~3~~.4 The Personalistic or Subjective Theory

Although Borel was perhaps the first to generate this theory as early as 1924, it was Ramsey who in 1931 first articulated the theory. It was later refined by de Finetti (1937 and 1974) and Savage (1954).

According to the theory, there is no such thing as an objective probability. Moreover, probability is a degree of belief for a given individual at a given time. Not only must the degree of belief be measured in some fashion, but also an individual's degrees of belief must conform to each other in a certain manner. The individual in question is an idealized one—in essence, one who behaves normatively.

Because the intensity of belief is extremely difficult to quantify, researchers elected to look at some property related to it. For example, Ramsey and de Finetti favored a behavioristic approach in which the degree of belief is expressed through the willingness to gamble. Thus, the probability of an event is the amount (say p) the individual is willing to bet, on a two-sided bet, in exchange for $1, should the event take place. By a two-sided bet we mean staking (1-p) in exchange for $1, should the event not take place.

The normative component of this theory lies in a feature known as “coherence.” Coherence ensures that the degrees of belief do not conflict (for example, it avoids a ~~“~~Dutch-Book~~” or “head I win, tails you lose” scenario~~). This is achieved by adhering to the calculus of probability.

4~~3~~.4.1 This Theory’s Virtues

According to this theory, probability is dictated by individual opinions, and thus “unknown probability,” “correct probability,” and “objective probability” cannot be achieved. To determine an individual’s probabilities, all a researcher need do is invoke the principle of indifference, apply a system of carefully conducted comparative wagers, or simply ask the individual. In this theory, any factor that an individual elects to consider is relevant and any coherent value is as good as another.

4~~3~~.4.2 This Theory’s Criticisms

The principal criticism leveled at this theory is that there can be little if any consistency in determining probability. For example, the theory has no provision to ensure that individuals with identical background information will declare identical probabilities. Given an individual’s action, it is difficult to separate the individual’s probabilities from his or her utilities. Because consistency is the hallmark of science, this theory is commonly refuted by scientists and engineers alike.

Perhaps the most important argument against this theory is that experiments by psychologists have shown that individuals do not declare probabilities that have coherence (i.e., they do not act according to the dictates of the calculus of probability).

A counter-argument to the above criticism is that the theory of personal probability is a normative one; it prescribes how we should act—not how we do act.

54.0 Which Interpretation of Probability is Appropriate for Reliability?

In the world of organizations such as the U.S. Government's Military Standards, automobile warranties, and commercial contracts, reliability analysis is entrenched in the relative frequency view of probability. With its claims of "objectivity," this position is reinforced by the peer review process for publication in many applied scientific journals.

This traditional interpretation to a degree has become outdated. For example, decision makers must make determinations under a number of intense restrictions, such as little or no testing (nuclear weapons, global climate change, and automotive reliability), one- or first-of-a-kind units (aerospace and medical applications), and economic testing (particularly in the manufacturing industry). Because of these and other restrictions, there has been a gradual shift toward the more personalistic view.

This shift can be traced back to the nuclear reactor industry (see WASH 1400, 1975) and evolved as the U.S. Government began to experience intense pressure over defense expenditures. As science and technology continued to evolve, other government and commercial areas began to feel the pressures of testing in complex and dynamic environments, particularly because such tests are either expensive, time consuming, or prohibitive for other reasons. The following are but five examples:

- Stricter emissions and performance requirements for automobiles while their engines operate at the cutting edge of physics and engineering (Kerscher et al., 2000).
- Determining software reliability by using complex and large computations (al-Mutairi et al., 1998).
- Using graphical, numerical, and simulation-based methods for a broad range of models found in reliability. This covers areas such as analyzing degradation data, in which failure is not dichotomous but continuous (Meeker and Escobar, 1998).

- Empirical techniques used to solve complex manufacturing techniques by using an empirical Bayesian approach to combine data with prior information (Samaniego and Neath, 1996).
- Using probability models for failure data analysis regarding maintenance and prediction related to a preliminary design. This is done by using influence diagrams, as well as a Bayesian approach (Barlow 1998).

From a philosophical standpoint, the personalistic interpretation of probability does not lead to the logical inconsistencies and other difficulties of communication mentioned before, nor does it demand the availability of a large amount of hard data (or preclude use if such data are available). This type of interpretation also enables us to do the following:

- Make statements of uncertainty about unique products or systems.
- Incorporate information for all sources that are deemed appropriate.
- Incorporate all relevant knowledge we have at any given time with the ability to update our probabilities (and hence reliabilities) as new knowledge becomes available.

~~A prime example in which the~~Th formal use of all relevant knowledge ~~could~~ can ~~have presented a different decision is the Challenger Space Shuttle tragedy. Instead of complete re~~reveal unanticipated problems before costly decisions are made, such as manufacturing recalls and disasters such as the Shuttle Challenger~~liance on the hard data from the solid rocket boosters (from a relative frequency view), a personalistic approach may have revealed the potential problems that subsequently led to the disaster~~. Therefore, we, feel that this is the point of view most appropriate for addressing the reliability problems in the 21st century.

From a pragmatic point of view, the dramatic evolution of our computational capabilities in recent years has made knowledge and information available in a variety of qualitative and quantitative forms. Large-scale simulations of complex, physical systems (such as transportation simulation, Beckman, 1997) provide gigabytes of information that must be analyzed and condensed in a form readily applicable for decision makers. Taking advantage of these information sources, including hard data, is what further motivates our point of view.

~~5.0 Probability: A Method to Quantify Uncertainty (Reliability)~~

~~This section provides an overview of several attempts at addressing the quantification of uncertainty. These attempts have contributors with diverse backgrounds, from philosophers and economists to physicists, psychologists, sociologists, engineers, statisticians, and mathematicians. Each approach has it own merits and flaws. Before we discuss these approaches, we must introduce some notation and terminology.~~

~~*5.1 Notation and Terminology*~~

~~Let E, E_1, . . ., E_i, . . ., denote several uncertain events of interest at some reference time, say τ. Although it is common to set $\tau=0$, it is important not to lose track of its presence. For example, E could denote an event that a deployed system accomplishes its mission. The complement of E is denoted by $\bar{E}$, the event that a deployed system fails to accomplish its mission (see Fig. 1). Another example could be that $E_i=\{T_i \geq t\}$, where T_i denotes the lifetime of the *i*-th sub-system of a deployed system (measured from the time of the system's deployment) and $\{T_i \geq t\}$ denotes the event that the *i*-th sub-system functions for at least t units of time. In this case, t is called the "mission time."~~

~~Let *H* denote the "history" or the "background information" available to the individual(s) contemplating the uncertain events, at time τ. In principle, *H* should encompass all that is known at time τ: scientific knowledge, engineering information, informed testimonies, design specification, physical models, computer codes, judgement, preferences, and hard historical~~

data on replicates of the uncertain event (if available). Thus at any time τ, there is the known *H*, and the unknown E, E_1, . . ., E_i, . . ., .

The fundamental problem of the treatment of uncertainty is how the uncertainty about E, E_1, . . ., E_i, . . ., at the τ, should be quantified in the light of *H*. To address this problem, several approaches have been proposed, some of which pay attention to the issue of “whose uncertainty” and others which impose restrictions on what *H* can and cannot contain. Some of these proposed approaches are as follows:

? probability,
? belief functions,
? possibility theory and fuzzy logic,
? upper and lower probabilities,
? Jeffrey's Rule of Combination,
? confidence limits,
? hypothesis testing with Type I and Type II errors,
? significance levels,
? maximum likelihood estimates, and
? goodness of fit tests.

Some of these approaches have a normative foundation, whereas others are *ad hoc*. We will focus on probability and make a case for it.

5.2 Probability and the Calculus of Probability

The calculus of probability consists of certain rules (or axioms) denoted by a number determined by $P^{\tau}(E; H)$, in which the probability of an event, E, is related to *H* at time τ. When the event E pertains to the ability to perform a certain function (e.g., survive a specified mission time), then $P^{\tau}(E; H)$ is known as the product’s reliability. Therefore, reliability is de facto the probability of a certain type of an event.

When the item in question is a human subject, the term "survival analysis," rather than reliability, is commonly used. As indicated above, the mission time need not be measured in units of time, but rather it can be taken from other performance metrics, such as miles traveled, rounds fired, cycles completed, or output produced.

The calculus of probability consists of the following three rules: convexity, addition, and multiplication. These rules are given mathematically in order:

? $0 \leq P^{\tau}(E; H) \leq 1$, for any event E;

? $P^{\tau}(E_1, \text{ or } E_2; H) = P^{\tau}(E_1; H) + P^{\tau}(E_2; H)$ for any two events E_1 and E_2 that are mutually exclusive—that is, they cannot simultaneously take place; and

? $P^{\tau}(E_1 \text{ and } E_2; H) = P^{\tau}(E_1 \mid E_2; H) \cdot P^{\tau}(E_2; H)$, where $P^{\tau}(E_1 \mid E_2; H)$ is a quantification via probability of the uncertainty about an event E (supposing that event E_2 has taken place).

The quantity $P^{\tau}(E_1 \mid E_2; H)$ is known as the "conditional probability" of E_1, given E_2. Note that conditional probabilities are in the subjunctive. In other words, the disposition of E_2 at time τ, were it to be known, would become a part of the history H at time τ.

The vertical line between E_1 and E_2 represents a supposition or assumption about the occurrence of E_2. Finally, $P^{\tau}(E_1 \text{ and } E_2; H)$ also can be written as $P^{\tau}(E_2 \mid E_1; H) \cdot P^{\tau}(E_1; H)$ because at time τ both E_1 and E_2 are uncertain events and one can contemplate the uncertainty about E_1 supposing that E_2 were to be true or vice versa.

The calculus of probability does not interpret probability. It does not do any of the following: it neither tells use what probability means, nor is it concerned with issues, such as the nature of uncertainty, whose uncertainty, how large should H be, and how to determine $P^{\tau}(E; H)$ and make the result operational.

What does the calculus do? It simply provides a set of rules by which the uncertainties about two or more vents combine or "cohere." Any set of rules for combining uncertainties that are in violation of the rules given above are said to be "incoherent" with respect to the calculus of probability. The next section discusses why these rules are necessary.

5.3 Why Subscribe to the Calculus of Probability?

In addition to the calculus of probability, there are a number of methods designed to comb and specify uncertainties, such as

? Jeffrey's rule of combination (Jeffrey, 1983),
? possibility theory and fuzzy logic (Zadeh, 1979),
? upper and lower probabilities (Smith, 1961), and
? belief functions (Dempster, 1968).

With so many theories, why should anyone subscribe to the calculus of probability? In a book by Howson and Urbach (1989), a number of contributors attempt to answer this very question; contributors range from gamblers and philosophers to mathematicians, decision theorists, behavioral scientists, and experts in artificial intelligence and knowledge acquisition. In the following subsections, we present some of the arguments used to justify the calculus of probability.

5.3.1 The Flaw of "Can't Win"

Since the time of Cardono (who lived during the 1500s), gamblers recognized that avoiding the rules of probability in games of chance resulted in a "sure loss" (i.e., a Dutch-Book). A classic example is a coin toss in which the scenario is as follows: heads you lose, tails I win.

5.3.2 The Scoring Rule

To justify the calculus of probability, de Finetti (cf. Lindley, 1982) used a "scoring rule," which is used to ask an individual assessing and uncertainty to declare a number that best describes said individual's uncertainty. Once the uncertainty reveals itself or is resolved, the individual is scored (i.e., rewarded or penalized) according to how close the declared number was to reality.

De Finetti's core argument is that under some very general conditions, an individual faced with a collection of uncertainties must use the calculus of probability to maximize an overall score. The above claim is true for a large class of scoring schemes.

5.3.3 Betting Coefficients

In horse racing, certain numbers known as "betting coefficients" are used. These numbers are odds on or against a particular horse or horses involved in a horse race. It has been demonstrated (for example, by Howson and Urbach, 1989) that to maximize winnings (i.e., determine the most accurate probability of success), the betting coefficients must follow the calculus of probability.

5.3.4 Behavioristic Axioms

The three previous subsections provided answers based on gambling and scoring scenarios; such scenarios could be objectionable to some individuals, particularly those who question the moral and ethical implication of gambling and fierce competition. Therefore, Ramsey (1931) and Savage (1972) proposed an alternative system of "behavioristic axioms" to justify the calculus of probability. Based in mathematics, the Ramsey-Savage argument is related to coherence and consistency (an excellent exposition of this argument is given by DeGroot, 1970).

~~This argument has two principal difficulties. The first is that the intuitive and natural elements of gambling and scoring are lost; axiomatic arguments tend to be abstract and therefore less appealing. The second and perhaps more serious difficulty is that behavioristic axioms are in actuality violated by (most) individuals (cf. Kahneman, et al., 1986). Despite these criticisms, behavioristic axioms prescribe normative behavior. It is not imperative that the calculus of probability be treated as axiomatic, but rather that it is seen as a consequence of certain behavioristic axioms, with the latter being taken as given (not withstanding some criticism).~~

66.0 Using Expert Testimony in Reliability Assessment

Once we adopt a personalistic interpretation of probability (reliability), we can assess reliability by using the calculus of probability on informed testimonies based on judgements, experience, simulations, or mathematical models. Expert testimony plays a crucial role particularly in cases in which hard data are unavailable or even impossible to obtain. In many instances, scientists and engineers have the knowledge and experience that can augment what little empirical evidence is available.

To maximize the accuracy of such expertise, it must be properly elicited and analyzed (cf. Meyer and Booker, 1991). Informed testimonies do not obviate the role of hard data when available. Instead, the personalistic view fuses the import of informed knowledge and hard data, the latter enhancing the former, via the calculus of probability and its extensions. For a detailed overview of this view, see Lindley and Singpurwalla (1986) and Singpurwalla (1988).

7.0 Closing Thoughts

As human science and technology continue to become more and more sophisticated, we will become more and more reliant on auxiliary information (especially because of the computer revolution) that augments direct hard data, which due to restrictions (such as political and societal constraints) may be scarce. Thus, the subjectivist view of probability can provide a paradigm for the quantification of uncertainty and information/data integration and therefore yield an accurate assessment of reliability. As a result, decision makers will have the best tools to apply to a new century of advanced science and technology and more sophisticated and complex societal and business issues.

Acknowledgements

Nozer D. Singpurwalla's work was supported in part by Grant DAAG-55-97-1-0323, The US Army Research Office, and by Subcontract No.35352-6085 with Cornell University under WO8333-04 with the Electric Power Research Institute and the US Army Research Office. Authors Booker, Bement, and Keller-McNulty would like to thank the Los Alamos Enhanced Surveillance Campaign, Technology Transfer Program office and Delphi Automotive Systems for their continuous support of this work.

References

1. D. alAl-Mutairi, Y. Chen, and N. Singpurwalla, "An Adaptive Concatenated Failure Rate Model for Software Reliability," *Journal of the American Statistical Association*, Vol. 93, No. 443, pp.1150-1163 (1998).

2. R. E. Barlow and F. Proschan, **Statistical Theory of Reliability and Life Testing**, Holt, Rinehart, and Winston, Inc., New York (1975).

3. R. E. Barlow, **Engineering Reliability**, SIAM (Society for Industrial and Applied Mathematics), Philadelphia, PA (1998).

4. R. J. Beckman, "TRansportation Analysis SIMulation System, The Dallas–Ft.Worth Case Study," TRANSIMS Release 1.0, Los Alamos National Laboratory Report LA-UR-97-4502, November 1997.

5. J. M. Booker and M. C. Bryson, "Decision Analysis in Project Management: An Overview," *IEEE Trans. on Engineering Management*, Vol. EM-32, No. l. (1985a).

6. J. M. Booker and M. C. Bryson, "Decision Analysis with Applications to Project Management: State of the Art Survey and Bibliography," *American Journal of Mathematical and Management Sciences*, Vol. 5, No. 1, pp. 001-062 (1985b).

7. B. de Finetti, "Foresight, Its Logical Flaws, Its Subjective Sources" (1937), in **Studies in Subjective Probability**, Kyburg and Smokler, eds., John Wiley and Sons, New York, pp. 93-158 (1964).

8. B. de Finetti, **Theory of Probability**, Vol. I, John Wiley and Sons, New York (1974).

9. M. DeGroot, **Optimal Statistical Decisions**, McGraw-Hill, New York (1970).

10. M. DeGroot, "Decision Theory," **Encyclopedia of Statistical Sciences**, Vol. 2, Kotz and Johnson, eds., John Wiley and Sons, New York, pp. 277-286 (1992).

11. P. Dempster, "A Generalization of Bayesian Inference (with discussion)," *Journal of the Royal Statistical Society* **B30**, pp. 205-247.

12. T. Fine, **Theories of Probability**, Academic Press, New York (1973).

13. G. Gigerenzer, Z. Swijtink, T. Porter, L. Daston, J. Bealty, and L. Kruger**, The Empire of Chance**, Cambridge University Press, Cambridge, MA (1989).

14. J. Good, **The Estimation of Probabilities**, MIT Press, Cambridge, MA (1965).

15. Hacking, **The Emergence of Probability**, Cambridge University Press, London and New York (1974).

16. Howson and P. Urbach, **Scientific Reasoning: The Bayesian Approach**, Open Court, La Salle, IL (1989).

17. R. C. Jeffrey, **The Logic of Decision**, Second Edition, University of Chicago Press, Chicago, IL (1983).

18. H. Jeffreys, **Theory of Probability**, Oxford University Press, New York (1961).

19. W. J. Kerscher, J. M. Booker, T. R. Bement, and M. A. Meyer, "Characterizing Reliability During a Project Development Program," European Safety and Reliability conference (ESREL) 2000, Edinburgh, Scotland, UK, May 22-24, 2000.

20. Kahneman, P. Slovic, and A Tversky, **Judgement Under Uncertainty: Heuristics and Biases**, Cambridge University Press, Cambridge, MA (1986).

21. V. Lindley, "Scoring Rules and the Inevitability of Probability," *International Statistical Review* **50**, pp. 1-26.

22. V. Lindley, **Making Decisions**, Second Edition, John Wiley and Sons, London, England (1985).

23. D. V. Lindley and N. D. Singpurwalla, "Reliability (and Fault Tree) Analysis Using Expert Opinion," *Journal of the American Statistical Association* **8**1, pp. 87-90 (1986).

24. N. R. Mann, R. E. Schafer, and N. D. Sinpurwalla**, Methods for Statistical Analysis of Reliability and Life Data**, Wiley, New York (1974).

25. L. E. Maistrov **Probability Theory: A Historical Sketch**, translated and edited by Samuel Kotz, Academic Press, New York (1974).

26. W. Q. Meeker and L. A. Escobar, **Statistical Methods for Reliability Data**, John Wiley and Sons, New York (1998).

27. M A. Meter and J. M. Booker**, Eliciting and Analyzing Expert Judgment: A Practical Guide**, Academic Press, London, England (1991).

28. M. A. Meyer, J. M. Booker, and T. R. Bement, "PREDICT—A New Approach to Product Development," 1999 R&D 100 Award winner, *R&D Magazine*, Vol. 41, p 161, September 1999 (also a Los Alamos National Laboratory document LALP-99-184, August 1999).

29. P. Ramsey, "Truth and Probability," **in The Logical Foundations of Mathematics and Other Essays**, Routledge and Kegan Paul, London, England (1934), reprinted in **Studies in Subjective Probability**, Kyburg and Smoklev, eds., John Wiley and Sons, New York (1964).

30. Reichenbach, **The Theory of Probability**, University of California Press (1949).

31. T. L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, New York (1980).

32. F. J. Samaniego and A. A. Neath, "How to be a Better Bayesian," *Journal of the American Statistical Association* **91**, pp. 733-742 (1996).

33. L. J. Savage, **The Foundations of Statistics**, Second Edition, Dover Publications, Inc (1972).

34. N. D. Singpurwalla, "An Interactive PC-Based Procedure for Reliability Assessment Incorporating Expert Opinion and Survival Data," *Journal of the American Statistical Association* **83**, pp. 43-51 (1988).

35. C. A. B. Smith, "Consistency in Statistical Inference and Decision (with discussion)," *Journal of the Royal Statistical Society* **B23**, pp. 1-37 (1961).

36. R. von Mises, **Probability, Statistics, and Truth**, Second Edition, George Allen and Unwin (1957).

37. WASH-1400, **Reactor Safety Study**, United States Nuclear Regulatory Commission, October 1975.

38. L. A. Zadeh, “Possibility Theory and Soft Data Analysis,” University of California, Berkeley, memorandum UCB/ERL M79/66 (1979).