Proceedings of the ASME 2018 Verification and Validation Symposium
VVS2018
May 16-18, 2018, Minneapolis, MN, USA

VVS2018-9308

MODELS, UNCERTAINTY, AND THE SANDIA V&V CHALLENGE PROBLEM

George A. Hazelrigg, Jr.
Independent Author
Vienna, Virginia, USA

Georgia-Ann Klutke
National Science Foundation
Alexandria, Virginia, USA

ABSTRACT

In this paper, we argue that the Sandia V&V Challenge Problem is ill-posed in that the answers sought do not, mathematically, exist. This effectively discredits both the methodologies applied to the problem and the results, regardless of the approach taken. We apply our arguments to show the types of mistakes present in the papers presented in J. of VVUQ along with the Challenge Problem. Further, we show that, when the problem is properly posed, both the applicable methodology and the solution techniques are easily drawn from the well-developed mathematics of probability and decision theory. The unfortunate aspect of the Challenge Problem as currently stated is that it leads to incorrect and inappropriate mathematical approaches that should be avoided and corrected in the current literature.

INTRODUCTION

Modern engineers are involved in the design, construction and operation of many large-scale systems for which physical testing is difficult, prohibitively expensive or politically unacceptable. These include civil infrastructure systems such as bridges, tunnels and dams, space systems such as the International Space Station, and national defense systems such as the nation's nuclear arsenal. For these systems, advanced digital computing provides the alternative of model-based approaches that enable analysis and virtual experimentation that would otherwise be impossible, too costly or simply prohibited. However, the use of model-based approaches must face questions of model verification, validity and uncertainty. As a result, a rather extensive literature is emerging around these topics.

To begin our consideration of model verification, validation and uncertainty quantification, it is imperative that we keep in mind that models merely support decision-making. Decisions are made by humans, not by models, and all real decisions are made under uncertainty. That is, the human decision-maker never knows with certainty the outcome of a decision he or she makes at the time of the decision. Unfortunately, these facts are not clearly recognized throughout the literature, and analytical methods are sometimes proposed as a surrogate to the human decision-maker. It follows that such methods may fail to have a rigorous mathematical basis. We propose to show here that the Sandia V&V Challenge Problem [1,2] is one such case.

Consider the problem posed by the mathematician Karl Weierstrass about 150 years ago: what is the largest positive integer? Our solution approach is as follows. Let n be a positive integer. Every positive integer has a square, n^2, which is also a positive integer. But, if n is the largest positive integer, it must be at least as large as n^2, namely $n \geq n^2$. This condition has one and only one solution, $n=1$. Ergo, 1 is the largest positive integer.

This "solution" is, of course, ridiculous, as 1 is the smallest positive integer. Yet the logic of the solution technique appears impeccable. What went wrong? The answer is that there is no largest positive integer—the solution to the problem posed simply does not exist. Hence, no solution technique, however logical it may seem, will find the correct solution because there is no correct solution.

It is important to comprehend the import of this example because of its implications for the Sandia V&V Challenge Problem.[1] The Challenge Problem consists of three parts:

[1] The Sandia Challenge Problem involves a fleet of 450 tanks that contain a mystery fluid and that have a design life of 20 years. Failure of a tank would constitute a serious problem. The Challenge Problem addresses the probability of failure of one or more tanks, P_{fail}, given limited test data.

- *Prediction*: the calculation of P_{fail} (the probability that at least one tank fails within its lifetime) and an uncertainty estimate on that prediction.
- *Credibility assessment*: the credibility of the estimate of P_{fail}, and
- *V&V strategy*: a logical and clearly defined strategy to gather evidence to demonstrate that the predictions are accurate.

As posed, it is implied that P_{fail} is a predictable physical property of the tank system, which has a "correct" value that can be calculated and upon which there is uncertainty. But a probability is not a physical property of a system; a probability is simply a belief [3,4]. Beliefs possess no physical properties: no mass, no charge, no temperature—in short, they cannot be measured, and they assume no unique value. As a belief, probabilities are never uncertain, they simply are what they are. Thus, it is mathematically incorrect to place a probability on a probability. And because there is no such thing as a probability of a probability, there is no mathematical provision for an uncertainty estimate on P_{fail}. As with the Weierstrass example, therefore, there is no correct answer and no correct solution technique to the determination of the uncertainty on P_{fail}. This leads us to the second part of the Challenge Problem, namely that there is no uniquely mathematically determinable credibility of any estimate of P_{fail} (this notwithstanding, it should be clear that an estimate would not be credible if the mathematics used to obtain the estimate are erroneous). To be sure, probabilities are beliefs that are held by individuals. So any "credibility" of the estimate of P_{fail} is simply a reflection of the credibility of the person who offers the estimate. Whatever the individual believes is, by definition, correct for that individual, and different individuals can and will have different values of P_{fail}. This holds true even in the case that the individual is a frequentist whose belief is that a previously obtained frequency is the probability of a future event. Finally, since P_{fail} is a belief and since beliefs are not predicted, it is actually irrelevant to consider demonstrating that the predictions are accurate.

As acknowledged in the Challenge Problem, the reason for estimation of P_{fail} is to support a decision that must be made under the condition of uncertainty [1], for example, to replace the tanks or not. Attempting to divorce the estimation of P_{fail} from this goal inherently renders the problem ill-posed and creates many ambiguities that confound proposed theories of probability and uncertainty estimation. The remainder of this paper will provide a review of probability theory in the context of decision making showing why there can be no uncertainty on a probability and hence why the Challenge Problem is ill-posed. We will then present a gedanken problem that will clearly illustrate the faults of the current approach to V&V. Finally, we will show how the way in which the Challenge Problem is posed leads to incorrect and inappropriate applications of mathematics and then point to the well-known mathematics that would address the Challenge Problem if it were properly posed.

RELEVANT FUNDAMENTALS OF PROBABILITY THEORY

Probability theory is a framework for thinking logically about the future. Although there have been other attempts at formalizing notions of uncertainty, to date probability theory provides the only mathematically sound platform to reason about future events. It provides a calculus for the analysis of beliefs, assuring that conclusions we may draw about uncertain events are consistent with a set of stated beliefs. Probability theory is different from statistics, which is a framework for thinking logically about the past. Statistics addresses the question, what conclusions are consistent with a given set of data? Since all data come from the past, statistics enables rational statements regarding the past. Beliefs can be altered by the introduction of data. Within the framework of probability theory, Bayes' Theorem provides a mechanism for altering beliefs based on data that assures the maximum preservation of information in the process.

It is important to understand at the outset that probabilities are beliefs [4] that belong to individuals. There is no requirement of consistency across a group of individuals, and there is no computation that combines individual beliefs into a "group" belief. Indeed, two individuals can hold divergent beliefs on the outcome of an event, even given identical information on the event, and yet we would say that both are correct. For example, in the case of a coin toss, Tom might believe that the coin is fair, while Lucy believes it is biased toward tails. It is then correct to say these are Tom's and Lucy's beliefs. There is no contradiction here, nor is there any way to combine these beliefs of Tom and Lucy. They simply do not share the same belief. Furthermore, Tom and Lucy may take opposing actions based on their different beliefs and, for each, they are taking the correct action. To interpret probabilities in any other way is inconsistent with probability theory, and it invalidates any conclusions one might draw from an analysis as well as the methodology by which the analysis is conducted.

The theory of probability centers around the concept of events [5]. An event is a phenomenon that is subject to randomness, but whose occurrence (or nonoccurrence) can be determined "after the fact," that is, once randomness has been realized. In other words, events convey observable information that is a manifestation of randomness. We are interested in quantifying our beliefs in the likelihood of the occurrence of events that are subject to the same random environment. In the case of a coin toss, we might be interested in two events, namely "the coin comes up heads" and "the coin comes up tails." In the case of the Challenge Problem, we might be interested in events corresponding to "exactly n tanks fail," where n is any non-negative integer. We use the notation $\mathcal{F}$ to refer to the collection (that is, set) of events of interest to us, and we refer to $\mathcal{F}$ as the "event space."

To assure consistency in the theory, the event space must have some structure. First, if we can observe that a particular event occurred, we should also be able to observe that the event has not occurred. (Were this not the case, for example, there would be no distinction between a tank failing and not failing.) This requires that, for any event, F, in the event space, its complement, $\overline{F}$, must also be in the event space. Second, if we can determine that an event F_1 occurred and an event F_2 occurred, we should also be able to determine if either F_1 or F_2 occurred. Hence, if F_1 and F_2 are in the event space, the union $F_1 \cup F_2$ must also be in the event space, and this condition must hold for countable unions as well. Third, we must be able to observe that something has occurred. That is, we construct an event, call it Ω, that carries no information other than "something happened;" we consider this a "certain" event. The structure thus placed on the event space comprises what is termed in mathematics a sigma field.

With these notions clearly defined, a probability measure P is a function that assigns a numerical value to each event in the event space such that:

- For any event F, $P(F) \geq 0$, and $P(\Omega)=1$;
- If F_1, F_2,… are mutually exclusive events (that is, if F_i occurs, then F_j cannot occur, for all $j \neq i$), $P(F_1 \cup F_2 \cup \ldots)=\sum_i P(F_i)$.

These are the Kolmogorov axioms of probability [6], and the above arguments provide the logic that underlies them.

An excellent illustration that justifies the need for a rigorous, logically consistent definition to quantify uncertainty, is provided by the Dutch Book Argument (DBA) [3,4,7,8,9]. The DBA centers around the Dutch, who were insuring shipping in the 19th century. They had discovered inconsistencies in the risk preferences of the shippers and formulated insurance packages that used these inconsistencies to their own advantage (the "Dutch Book"). The DBA was formalized by Ramsey [10] and de Finetti [11,12] beginning in the 1920's and, using mathematics no more complex than high school algebra, provides a robust rationale for the axiomatic development of probability by Kolmogorov. Furthermore, it can be shown that no other approach to thinking about the future is consistent with the notion that a gambler will not accept a wager that is a certain loss; that is, the Kolmogorov axioms are both necessary and sufficient to ensure coherency.

A probability is a "measure" in the mathematical sense only (that is, it is a countably additive set function). It has no objectively measurable physical properties. Mathematics does not prescribe how one is to establish a probability measure; one is free to use models, available data, a divining rod, or three weyward Sisters, if desired (perhaps Macbeth [13] should have done a better job of validating his model). A decision maker chooses the (valid) probability measure that best reflects his or her beliefs about the uncertainty surrounding the given decision scenario. And as a decision-maker's beliefs may change (for example, because of consideration of additional evidence), his or her probability measure may also change before a decision is made. Because different decision makers have different beliefs about uncertainty, there is no, nor can there be any, "correct" or "incorrect", or even "better" or "worse" probability measure, provided that the probability obeys the rules described above. Neither can there exist any "uncertainty" related to a probability, as a probability itself encodes the decision-maker's uncertainty about all events of interest. It is mathematically incorrect to attempt to assign a "probability" (or an "uncertainty estimate") to a probability, because the object of the desired assignment is not a set function. It follows that, in the context of the Challenge Problem, there is no provision in probability theory for the existence of any uncertainty measure on P_{fail}.

As mentioned earlier, probability theory's operational calculus provides all of the mathematical tools necessary to draw conclusions about events once the probability measure has been chosen by the decision maker. Indeed, in many cases the decision maker has only to choose the probabilities for a foundational subset of all possible events in order to "complete" the measure for other events in a consistent fashion. The operational calculus includes Bayesian analysis, which provides a mathematically sound mechanism for updating a probability measure based on empirical evidence in the form of data.

Importantly, the Kolmogorov axioms, which preserve countable additivity, also preserve the cardinal number sets. This is necessary if one chooses to perform a numerical analysis of probabilities, as it is not mathematically permissible to merge logic systems whose axiomatic bases are contradictory. Alternative theories of uncertainty—for example, Fuzzy Logic, probability bounds and Dempster-Shafer theory—do not preserve countable additivity and, at least in the case of Fuzzy Logic, the cardinality of numbers with the result that it is mathematically incorrect to perform a numerical analysis in the context of this logic system.

DESCRIPTIVE AND PREDICTIVE MODELING IN ENGINEERING

Descriptive models encompass those models that represent our extant understanding of relationships between physical properties of a system. They include, for example, "laws of nature" such as Hooke's law and Newton's laws, and they may describe causal relationships. It is possible to design experiments that provide objective validation of these models. Predictive models, on the other hand, provide a probability law on the outcome of an event. For example, a predictive model of a coin flip might identify the outcomes head and tails each with probability 0.5. Predictive models play a key role in the support of engineering decision making and, particularly, in engineering design. While descriptive models are often (incorrectly) used to support predictions, they themselves do not provide a probability law on the outcome of an event, and their use may lead to poor decision making.

Predictive models are often built upon the understanding obtained from descriptive models. And, while a predictive model may appear to be very similar to the descriptive model upon which it is based, the two are different models. While descriptive models can be objectively validated via physical testing, it is not possible to objectively validate a predictive model. Consider the predictive model of the outcome of a coin flip. One may think that this model can be validated by flipping the coin many times and using the frequency of heads and tails to validate the probabilistic prediction of the model. However, this procedure relies on the subjective assumptions that the flips are independent and identically distributed (IID), and that the future behavior of the coin will be the same as its past behavior. These assumptions cannot be tested.

Why do we create predictive models? In a problem such as that of the tanks posed in the Sandia Challenge Problem, it is absolutely the case that one could make an ad hoc judgment regarding the likelihood of a tank to fail, even in the total absence of any physical evidence. However, in the Challenge Problem case, we are given a belief that the tank will fail if the Von Mises stress exceeds the yield stress at any point on the surface of a tank. We also believe that stress and strain are proportional to each other. The reason that we create a mathematical model is because we believe that we are less likely to come to a conclusion that is contrary to these beliefs using the model to enforce them than to simply use our ad hoc judgment. Uncertainty arises because we also believe that there are considerations that we are not taking into account in enforcing the above beliefs and in the belief that there is unaccounted-for variability in such things as dimensions, tolerances, material properties, and so on. Thus, not only are probabilities beliefs, but predictive models that we might construct to assist in our estimation of probabilities are themselves constructs based on beliefs. The purpose of modeling is, therefore, to enforce consistency between our underlying beliefs and the conclusions upon which we will take action.

A GEDANKEN EXAMPLE

A gambler is going into a casino to gamble. Indeed, engineers *are* gamblers, placing bets on the outcomes of their professional decisions. To keep this example simple, we will say that the gambler intends to bet on the outcomes of a series of coin flips. As she is about to enter the casino, a person approaches her and says, "I am an engineer, and I write predictive models." This is what engineers often do. "I have written a model to predict the outcome of a coin flip." The gambler expresses her skepticism that such a model would be valid. The engineer suggests that she borrow the model and validate it. So the gambler does a series of tests. She runs the model, gets a prediction and then flips the coin. She does this very many times, and she notes that every time the model predicts heads, the coin lands tails, and every time the model predicts tails, the coin lands heads. We now pose two questions. (1) Should the gambler buy the model? (2) Is it valid?

With respect to the first question, it is obvious that the gambler could bet against the result provided by the model and win every time. It doesn't get better than this. So, clearly, it would be to the advantage of the gambler to have the model. But is the model valid? The results provided by the model itself could not be further from an accurate representation of the physical reality. By the standards of the V&V Challenge Problem, we would have to say that the model is not valid. Yet, this makes no sense. Why would we choose to say that a model that enables perfect decision making is not valid? In fact, one definition of "validity" is "producing the desired results; efficacious" [14]. This model provides perfect results for the task at hand and, by this definition, must be valid.

The resolution of this paradox gives us considerable insight into the question of model validity. To begin, we must recognize that the immediate output of a model—a result on a computer screen or a printout, for example—is never used to make a decision. A decision is made by a person, and the information used in making the decision is, of necessity, in the head of that person. Thus, for a model result to be used in a decision, it must get from the model output into the head of the decision-maker—it must transition from output to belief. In the process of making this transition, the decision-maker "updates" the model in accordance with her beliefs about such things as the accuracy of the model. If done optimally, that is, to preserve the information in the best possible manner, the update would be in accordance with Bayes' Theorem and, in the example presented, it would be logical to flip the output: heads becomes tails and tails becomes heads. When this occurs, the belief in the decision-maker's head is obviously valid.

This example clearly shows the need to validate the decision-maker's beliefs as opposed to the model output. It emphasizes that probabilities inure to a decision-maker. And, since different decision-makers will hold different beliefs, probabilities are strictly personal, and model validation is a personal thing. A model that is not at all valid to one decision-maker can be perfectly valid to another decision-maker in an otherwise identical decision situation. As a result, it makes no sense to question the validity of a model in the abstract, that is, in the absence of a specific decision situation and for a specific decision-maker. Yet, this is precisely what the Sandia V&V Challenge Problem seeks to do.

ERRORS COMMITTED IN SOLUTIONS TO THE V&V CHALLENGE PROGLEM

In this section, we comment briefly on some of the shortcomings in the description of the Challenge Problem that lead to errors in the papers included in J. VVUQ, Vol. 1 [15-20].

- The Challenge Problem asks for a "computation" of P_{fail}, and an uncertainty estimate on that prediction. This clearly suggests that probabilities are a physical property of a system, that there exists a "correct" P_{fail} that one might, in theory, compute, and that there would be uncertainty on any actual computation of P_{fail}. Probabilities are subjective—they are beliefs held by a decision-maker and thus are, by definition, correct statements of the beliefs of the decision maker. Accordingly, there is no single "correct" estimation of

P_{fail}, and there is no uncertainty on an estimation of P_{fail}. It follows that no method that purports to compute a unique "correct" value for P_{fail} can be correct. Further, there is no provision in probability theory for uncertainties on probabilities. Thus, an analysis of the uncertainty on P_{fail} constitutes an attempt to place a numerical estimate on something that does not exist, and accordingly, whatever the supporting analysis may be, it is of necessity incorrect.

- The Challenge Problem, as stated, encourages analysts to make use of ad hoc approaches to the estimation of P_{fail}, such as the use of fuzzy probabilities [21], interval probabilities and Dempster-Shafer structures. These methods are thoroughly debunked by Bernardo and Smith, and Haack [7,22,23], with concurrence from the entire mathematical community, and should not be used. Many of these methods fail to satisfy countable additivity, and are thus vulnerable to the Dutch book and contradictory to the axioms that form the basis of the cardinal number sets.
- The Challenge Problem is not specific about validation of the solution methodology. Mathematically, validation of a process or model must be based on a set of clearly specified axioms, assumptions and beliefs, which form the basis of the logic system that underlies the process or model and define its boundaries or limitations. Moreover, the axioms from which a mathematical method is derived must constitute a self-consistent set [24]. That is, a logic system cannot be based on a set of self-contradictory axioms, and the validity of any mathematical method or analysis rests on its derivation. The Challenge Problem does not ask that solutions be based on methodologies derived from clearly stated axioms, beliefs and assumptions nor does it request that these be explicitly identified. While this would normally preclude an assessment of the validity of the models and approaches taken, as the Challenge Problem is ill-posed such that we know a solution does not exist, we can conclude that all approaches to its solution must, therefore, be non-valid.
- In asking for a computation of P_{fail} without reference to a specific decision context, the Challenge Problem fails to recognize that probabilities belong to individuals and are of value to a decision maker only if they are the belief of the decision maker. In the likely case that the beliefs of the decision maker do not concur with those of the analyst, the resulting computation of P_{fail} will be of little or no value to the decision maker. One might argue that a frequentist would be willing to accept as his/her belief a computation of P_{fail} that is based on frequency data, but this argument restricts the utility of the result to decision makers who are frequentists and also relies on the assumptions that these decision makers believe all models used in the computation of P_{fail} and the validity of the data upon which the frequencies are derived (for example, that the experiments from which the data derive are IID).
- The Challenge Problem encourages an economic cost-benefit analysis, which is not the appropriate approach for analysis of the value of model validation. To begin, the validity of a model is personal, that is, it pertains only to a specific decision-maker, and the value of a model depends on the risk preferences of the decision-maker, not merely on expected values. Model validation activities provide evidence that a decision-maker can use to modify his/her prior beliefs. The correct approach to the economic evaluation of the value of such information is given by value-of-information theory [25-27], which is a subset of decision theory. In this theory, choices are made on the basis of expected utility, and the value of the information relates to the extent that the information can change decisions.

ALTERNATIVE DEFINITIONS OF PROBABILITY

The classical definition of "probability" is consistent with the Kolmogorov axioms. These axioms lead to the interpretation of probability as a belief, which cannot be uncertain. In the case of Kolmogorov probability, a rigorous mathematical framework for decision making has been derived [28], and it is entirely consistent with the accepted mathematics of number theory, arithmetic, calculus, algebra, and so on. In mathematics, this framework is referred to as decision theory. In economics, it is more commonly referred to as utility theory. Alternative theories that allow for uncertainty in the numerical value of a probability are not consistent with the mathematics of decision theory and, in some cases such as Fuzzy Logic, even number theory or arithmetic. Indeed, the axioms of Fuzzy Logic contradict the axioms of number theory. As a result, it is mathematically inconsistent to use Fuzzy Logic and arithmetic in the same analysis, and doing so will lead to results that have no conventional interpretation. Equally important, to our knowledge, for none of the alternative definitions of probability has there been derived a mathematically rigorous alternative to classical decision theory. It is certainly the case that all of the papers presented in VV&UQ, Volume 1, fail even to acknowledge the need for such a rigorously derived decision framework. Yet, the need for such a framework should be intuitively obvious: if one changes the definition of a word such as "probability," as has been done in the papers of this journal [15-20], one should no longer use the word as though it remains as previously defined.

RESTATING THE V&V CHALLENGE PROBLEM

The physical description of the Sandia V&V Challenge Problem is a sound basis for illustration of V&V. Where the Problem goes astray is in the questions that it poses, beginning with the calculation of P_{fail} and the uncertainty on this calculation. If instead, we reformulate the problem asking the question—*In light of the test and other available data, would you replace the tanks?*—then the problem is well-posed, and the solution approach is clearly defined by extant mathematical techniques including probability theory, von Neumann-Morgenstern utility theory, game theory and social choice theory [29]. The choice appropriate to a specific decision-maker is that alternative that has the greatest expected utility [3]. However, while subject to a reasonable set of axioms this is provably the only correct approach, one should note that there is no one-and-only "correct" solution to the question.

To begin, we can model the physics of the tanks and relate stress and strain to internal pressure and to each other. This modeling can be by any of several relevant models, from simple shell models to finite element models as suggested in the Challenge Problem. Uncertainty, dependent upon which modeling technique is used, will exist on the computation of these quantities, and it is straight forward to estimate that uncertainty. Next, using Bayes' Theorem we can update the computed values (and perhaps even the model) given data from various tests, and the (pseudo) decision-maker can internalize these results based on his/her personal beliefs. Then, classical decision theory [3,28,30] can be applied to enable the decision-maker to conclude a decision that is entirely consistent with his/her beliefs, preferences and available alternatives. Finally, should it be desired to determine the economic value of the analysis, this may be determined also relying on the classical theory and to the subset of that theory referred to as "value of information" [3,25]. Thus, it is seen that the extant classical theory is fully applicable to the case of the Sandia V&V Challenge Problem, and that this theory is adequate for the complete solution of this problem when the Problem is well-posed.

Verification of the solution approach may also be accomplished via mathematical proof such as that provided by Hazelrigg [3]. Within the theory, definition of terms is precise. For example, terms such as "probability" and "preference" are quite precisely defined. Further, the theory is built on defensible axioms, which not only offer a proof of the validity of the approach, but simultaneously show that no alternative theory is valid, and they show that our theory is fully compatible with other mathematics including number theory and arithmetic. Decision theory has evolved over the past 300 years, with contributions from such great mathematicians as Bernoulli [31], Dodgson[2] (Lewis Carroll) [32,33] and von Neumann [28]. The theory has been thoroughly vetted over the past half century, and its validity is accepted without question by all who truly understand it.

CONCLUSIONS

The Sandia V&V Challenge Problem is ill-posed. As stated, there is mathematically no answer to the Problem and, hence, no correct answer. Ergo any attempt to find an answer will obtain an incorrect result via an irrelevant methodology. The question of relevance is whether to replace the tanks or not. Posed in this manner, straight-forward application of the mathematics of probability and decision theory leads directly to the desired result. However, this result is personal, that is, it is not necessarily the same for different decision-makers who will, in general, hold different beliefs and preferences (including risk preferences).

It is also the case that the validity of the decision approach rests on the validity of the mathematics used. The extensive mathematics literature on prediction and decision making is quite clear that the range of valid mathematical approaches is rather limited. To be specific, Kolmogorov probability theory is widely accepted by mathematicians as the only self-consistent approach to prediction. Indeed, the generally accepted concept of a quantity called "probability" is defined by the Kolmogorov axioms and only by these axioms. The concept will not remain the same in other axiomatic systems. More specifically, ad hoc methods such as Fuzzy Logic, probability bounds, Dempster-Shafer theory, and related methods have significant faults and should not be used [7,22,23]. Classical decision theory utilizing von Neumann-Morgenstern utility is provably the only self-consistent approach to the analysis of decisions. However, it must be realized that decisions are made only by individuals, not by groups. Groups have emergent behaviors that are properly analyzed by the mathematics of game theory. Further, it is a mistake to assign a preference function to a group [29], and members of a group will inevitably have differing beliefs.

Probability theory rests on a very solid axiomatic base that extends well into set theory, measure theory and number theory. Applied correctly, it provides a calculus for the manipulation of beliefs, leading to conclusions that are consistent with a set of underlying beliefs. However, applied incorrectly, one can place no credibility at all in the results. Ad hoc redefinition of "probability" is mathematically incorrect, and it will destroy the credibility of any result. Probabilities are always subjective; adjectives such as objective, aleatory and epistemic simply do not apply, and they are misleading[3]. Probabilities simply are, they are not predicted, and they are not uncertain.

Finally, it is very important that publications in the field of V&V be adequately vetted to assure that they are based on a rigorous mathematical foundation. Failure to assure the quality of such publications creates an environment of inappropriate theories that result in a counterculture to that of the rigorous mathematics, thereby confounding and delaying actual progress in the field. We suggest that the Sandia V&V Challenge Problem be restated in a way that is well-posed, and that the V&V research community be challenged to apply the well-

[2] The encounter between Alice and the Cheshire Cat presents the underlying axioms of modern, normative decision theory.

[3] The apparent need for distinctions such as aleatory and epistemic evolves from an attempt to derive a theory of probability estimation independent of a decision context. When probabilities are estimated and analyzed only in the context of a specific decision, the ambiguities that result in such distinctions disappear, and the terms themselves become clearly irrelevant.

known and well-vetted appropriate mathematical techniques to the solution of the problem.

DISLAIMER

The views expressed in this paper are strictly those of the authors and not necessarily those of the National Science Foundation or the Federal government.

REFERENCES

[1] Hu, K., Carnes, B., Romero, V., *Guest Editorial*, “Introduction: The 2014 Sandia Verification and Validation Challenge Workshop,” J. VV&UQ, Vol.1, March, 2016.

[2] Hu, K. and Orient, G.E., “The 2014 Sandia Verification and Validation Challenge Problem Statement,” J. VV&UQ, Vol.1, March, 2016.

[3] Hazelrigg, G.A., “Fundamentals of Decision Making for Engineering Design and Systems Engineering,” www.engineeringdeicisionmaking.com, 2012.

[4] Caves, C.M., Fuchs, C.A. and Schack, Rüdiger, “Quantum probabilities as Bayesian probabilities,” arXiv:quant-ph/0106133v2, Nov. 14, 2001.

[5] Ash, R.B., “Basic Probability Theory,” Dover Books on Mathematics, Mineola, NY, 2008.

[6] Kolmogorov, A., “*Grundbegriffe der Wahrscheinlichkeitsrechnung,” Springer, 1956; translation: “Foundations of the Theory of Probability,” Chelsea, New York, 1956.*

[7] Bernardo, J.M. and Smith, A.F.M., “Bayesian Theory,” John Wiley & Sons, Ltd, New York, 2000.

[8] Hájek, A., “Scotching Dutch Books,” *Philosophical Perspectives*, 19, issue on Epistemology, ed. John Hawthorne, 2005.

[9] Hajek, A., “Dutch Book Arguments,” Chapter 7 in The Oxford Handbook of Rational and Social Choice, ed. Paul Anand, Prasanta Pattanaik, and Clemens Puppe, pp. 173-195, Oxford University Press, 2009.

[10] F.P. Ramsey (1926) "Truth and Probability", in Ramsey, 1931, The Foundations of Mathematics and other Logical Essays , Ch. VII, p.156-198, edited by R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company.

[11] de Finetti, B., “Theory of Probability,” John Wiley & Sons, Ltd, New York, 1990.

[12] de *Finetti, B., “Sul significato soggettivo della probabilità,” Fundamenta Mathematicae.* ***17****, 1931, pp. 298– 329.*

[13] Shakespeare, W., “The Tragedy of Macbeth,” 1606.

[14] *Houghton Mifflin Dictionary.*

[15] Beghini, L.L. and Hough, P.D., “Sandia Verification and Validation Challenge Problem: A PCMM-Based Approach to Assessing Prediction Credibility,” J. VV&UQ, Vol.1, March, 2016.

[16] Choudhary, A., Voyles, I.T., Roy, C.J., Oberkampf, W.L. and Patil, M., “Probability Bounds analysis Applied to the Sandia Verification and Validation Challenge Problem,” J. VV&UQ, Vol.1, March, 2016.

[17] Li, W., Chen, S., Jiang, Z., Apley, D.W., Lu, Z. and Chen, W., “Integrating Bayesian Calibration, Bias Correction, and Machine Learning for the 2014 Sandia Verification and Validation Challenge Problem,” J. VV&UQ, Vol.1, March, 2016.

[18] Xi, Z. and Yang, R-J., “Reliability Analysis With Model Uncertainty Coupling With Parameter and Experiment Uncertainties: A Case Study of 2014 Verification and Validation Challenge Problem,’ J. VV&UQ, Vol.1, March, 2016.

[19] Mullins, J. and Mahadevan, S., “Bayesian Uncertainty Integration for Model Calibration, Validation, and Prediction,” J. VV&UQ, Vol.1, March, 2016.

[20] Paez, P.J., Paez, T.L. and Hasselman, T.K., “Economic Analysis of Model Validation for a Challenge Problem,’ J. VV&UQ, Vol.1, March, 2016.

[21] Zadeh, L., “Fuzzy Logic for the Management of Uncertainty,” John Wiley & Sons, 1992.

[22] Haack, S., “Deviant Logic, Fuzzy Logic: Beyond the Formalism,” The University of Chicago Press, Chicago, 1996.

[23] Haack, S., “Philosophy of Logics,” Cambridge University Press, Cambridge, UK, 1978.

[24] *Hofstadter, D.R., “Gödel, Escher, Bach,” Basic Book, New York, 1979.*

[25] *Howard, R.A., “Information Value Theory,” IEEE Transactions on Systems Science and Cybernetics (SSC-2), 1966, pp. 22-26.*

[26] *de Neufville, R., “Applied Systems Analysis: Engineering Planning and Technology Management,” McGraw-Hill, New York, 1990.*

[27] *Hazelrigg, G.A.,* “Economic Viability of Pursuing a Space Power Concept,” Journal of Energy, Vol. 1, No. 2, March-April 1977, pp. 93-99.

[28] von Neumann, J. and Morgenstern, O., “The Theory of Games and Economic Behavior,” Princeton University Press, Princeton, NJ, 3rd edition, 1953.

[29] Arrow, K.J., Social Choice and Individual Values, John Wiley & Sons, 2nd edition, 1963.

[30] Tribus, M., “Rational Descriptions, Decisions and Designs,” Pergamon Press Inc, 1969.

[31] Bernoulli, D., “Exposition of a New Theory on the Measurement of Risk” (1738), translation in A.N. Page, editor, *Utility Theory: A Book of Readings*, John Wiley & Sons, Inc., New York, 1968, pp. 199-214.

[32] Dodgson, C.L. (Lewis Carroll), “Alice in Wonderland,” 1865.

[33] Hazelrigg, G.A., “The Cheshire Cat on Engineering Design,” Quality and Reliability Engineering International Vol. 25, 2009, pp. 759-769.