An Evidence-Based Approach to Fidelity

VVSIMLAB CONCEPTS

D. E. STEVENSON

Abstract.

(1) Usurp the mechanisms of HOL-Light.
(2) Introduce the following
   (a) General Systems Theory
   (b) Formal Methods
   (c) Conventional-Interval-Fuzzy mathematics
   (d) Simulation Generation
(3) For each concept, introduce
   (a) create that puts new thing in dictionary
   (b) mk that constructs a representation
   (c) dest that pulls out the data
   (d) is that answers questions about concept
(4) FM ideas are basically Hoare triples as connotational semantics and $\lambda$-calculus as denotational semantics.
(5) Get the background from the IMSA 105 book. Change to axioms.
(6) Concept is that components are containers of information, so there is an association of logic to FM and FM to GST. It’s all about semantics.
(7) ST, CC, Beh triangle come from dynamical systems. Dynamical systems is to GST as
calculus is to kinematics is to dynamics.

(8) Encode GST into Hoare triples as logical dynamical systems.

(9) The deeper logical systems that you used in to SCSC paper in language and in evidence

(10) OK when ST/CC and proofs run the simulation.

(11) Category notation. Are they categories? We have GST cat formalization possible. John
Baez does a lot; guess they like it. Enforces algebra as foundation.

(12) Evidence!!!!!

(13) Trick is to get your arms around whole process.

(14) Find the Feynman quote on “we only use math when we don’t know what we’re doing.”

For the FOPL, you need the FOPL logic, the interface logic, and the application. See below.

Hoare triples as connotative semantics

It’s about the semantics, stupid. It’s also

about categories as functions, not a foundations area.

The ST-CC-Beh triangle is everything!

Rules as basis: even logic is schematic: pattern-¿substitution-¿rewrite.

Four pieces of schema: (1) pattern recognition via unification; (2) condition/criteria satisfaction;
(3) plan generation by substitution-rewrite; and (4) implementation by substitution-rewrite.

Abstract. The Foundations 2004 Workshop focused on risk assessment and risk mitigation as
yey relate to verification, validation, and accreditation. Risk and uncertainty are primarily due to
ack of information during decision-making. Therefore, we take decision-making as the context for
fification, validation, and accreditation. The state of the art of decision-making is reviewed. The
esearch of Klir and Weirman relating to information and information theories is reviewed, which
eads to reviews of uncertainty and risk analysis. The material is illustrated using a simple modeling
nd simulation example. Recommendations are presented on incorporating decision-making into
fification, validation, and accreditation processes.

Keywords: Decision-making, fuzzy arithmetic, interval arithmetic, information.

1. Introduction

The central theme of the Foundations 2004 Workshop on Verification, Validation, and Ac-
creditation (Foundations 2004), held at Arizona State University on 13–15 October 2004 [?], was
the role of uncertainty and risk in verification, validation, and accreditation (VV&A). Unlike Foundations 2002, which documented the state of the art in VV&A, Foundations 2004 was a working research meeting with broad representation from academia, Department of Defense (DOD), Department of Energy (DOE) National Laboratories as well as significant foreign participation. The majority of the presentations explored risk, risk assessment, uncertainty and uncertainty quantification. These presentations approached the subject from a standpoint requiring significant statistical expertise; however, many, if not most, VV&A decisions are made by non-statisticians. Hence, there appears to be an education requirement if statistical VV&A processes are to be integrated into system development because of the levels of expertise needed to interpret results. One approach to such education is to adopt a decision-making model because risk and uncertainty are naturally addressed in that paradigm. A decision-making approach to VV&A can be supported by developing computational libraries for uncertain systems theory, probability, and alternative arithmetics.

The next section reviews the state of the art in decision-making, while Section 3 considers the long history of psychology and mathematics in support of decision-making. Section 4 introduces work by Klir and Weirman relating to information and information theories, and Section 5 reviews uncertainty and risk principles. Section 6 relates the material to the modeling and simulation example. Finally, Section 7 presents the recommendations.

2. VV&A and Decision-Making

While VV&A is often defined in terms of questions to be answered, an alternative approach is to consider it as decision-making processes, allowing for the incorporation of four centuries of research. The most important issues regarding decisions are the frame of reference and the inclusion of the decision-maker’s value sets.

Assumptions. Decision-making processes are carried out by rational decision-makers. In the modeling and simulation (M&S) context, there are often multiple decision-makers who must share a single frame of reference (frame). This frame of reference, which is primarily terminological, serves as a context for the definition values needed to specify a context. In addition, it is assumed that the decision must be made in an uncertain environment.

Frame of Reference. Of all VV&A process components, the frame of reference is the least discussed, albeit perhaps the most important. The frame of reference, or “world view,” involves the implicit assumptions, methods, metaphors, among others, that go into a VV&A decision, providing
the context in which it is made. At Foundations 2004, it was evident that there were at least two
frames of reference: the Department of Defense and the Department of Energy. Kilikauskas and
Hall [?] described the DOD frame, and papers by Logan and Nitta [?] plus the paper by Pilch,
Trucano, and Oberkampf [?] described two frames of reference of the National Laboratories. The
focus of the DOE workers at the Workshop was, and still is, the nuclear weapons stockpile, primarily
a computational science and engineering (CSE) problem, a term denoting modeling and simulation
typified by complex physical models. CSE VV&A has an entirely different frame of reference than
many DOD VV&A efforts, including fundamental differences in the definitions of basic terms.

Methodological Assumptions. Since the decision-makers seek to clarify the situation as
best they can, it is assumed that they will have goals and objectives that are part of the decision
and that they will be able to measure their “levels of satisfaction.” These levels of satisfaction are
measure-based values in a particular frame of reference.

3. Decisions

Research into the subject of decisions under uncertain circumstances has a long history, the
earliest example being Blaise Pascal and Pierre de Fermat, among others, who, in 1654, considered
decisions that arise in the gambling context [?]. Jacob (James) Bernoulli is generally considered
the founder of statistics with his Ars Conjectandi (1713), while Pierre Simon de LaPlace in Théorie
Analytique des Probabilités (1812) was the first to consider systematically probability and statistics
in science and in other practical problems. While mathematics can help justify a decision, decision-
making is essentially a human activity and, hence, primarily psychological in nature. As a result,
decision-making is inherently a modeling exercise; indeed, many non-CSE models arise from the
need for more information in the decision-making process.

It has long been established in psychology [?] that human performance in probabilistic
inference is suspect. In 1982, Kahneman et al. [?] proposed a unified, problem-solving view of
decision-making based on John Dewey’s problem-solving paradigm:

(1) Clearly state the problem using the proper vocabulary.
(2) Establish goals, including the constraints and the criteria.
(3) Gather relevant information, clearly listing what is known and what is not, addressing cases
and/or experiments to decide.
(4) Generate possible solutions.
(5) Apply the constraints and criteria by conducting thought-experiments or developing prototypes; develop more tests and experiments.

(6) Synthesize a solution, testing it as completely as possible.

(7) Evaluate and test; review.

This systemization has led to a new understanding of classical decision-making that assumes infinite time and resources, leading Zsambok and Klein [?] to investigate decision-making in the context of short time frames, now known as naturalistic decision-making.

Zsambok et al. list six parameters for decision-making:

(1) Situational awareness
(2) Option evaluation
(3) Fluidity of the situation
(4) Ambiguity of information
(5) Stability of goals
(6) Time constraints
(7) Previous experience

While these are sufficient to categorize common “decision theory” approaches, many standard engineering practices, such as reliability analysis, are decision-making approaches.

4. INFORMATION

Fundamental to decision-making are the issues of information, uncertainty, and risk. Lack of information leads to uncertainty. In this context, information is not Shannon information, but information as a measure of the ability to distinguish among states. To describe these states, a model of information is needed. Klir and Weirman [?, ?] were the first to describe alternative information models in relation to the alternative forms of uncertainty quantification.

An information model is defined as a model in the decision-makers’ frame of reference relating the decision model to the uncertainty and risks.

5. UNCERTAINTY AND RISK

VV&A are the decision processes for justifying the underlying model in a specific frame, meaning deciding whether or not the model or simulation fits into the knowledge represented within the frame. Uncertainty is an inevitable outcome of attempting to justify new knowledge
because some accepted knowledge in the frame may be at odds or even contradictory to that new
knowledge being introduced. Rhetorically, it might be asked which warrants for knowledge are
correct and how much of the old knowledge needs to be preserved.

Although Foundations 2004 focused on epistemic (lack of knowledge) and aleatory (fundamental variability) uncertainty, there are others. The judgment literature describes informational uncertainty as the inability to distinguish alternatives from one another. Fuzzy logic introduces lexical uncertainty brought on by inherent vagueness in language.

To develop information models, methods for identifying uncertainty and then quantifying it must be developed. Klir and Weirman describe several such models of uncertainty quantification [?]

- **Bayesian.** The Bayesian (or subjective belief) interpretation of probability is that it measures strength of belief.

- **Frequentist.** The frequentist interpretation of probability is that it measures chance as the long-run percentage of occurrence of events.

- **Fuzzy.** Fuzzy logic is based on the concept that set membership is not “all or nothing.” The fuzzy community has developed mathematical theories through measure theory with classical theories as subtheories. Fuzzy measure theory includes standard measure theory, which supports axiomatic probability.

- **Possibility Theory.** Possibility theory allows reasoning to be used on imprecise or vague knowledge, making it possible to deal with uncertainties on based this new knowledge.

- **Imprecise Probability.** Imprecise probability is a generic term for the many mathematical or statistical models which measure chance or uncertainty without sharp numerical probabilities. These models include belief functions, Choquet capacities, comparative probability orderings, convex sets of probability measures, fuzzy measures, interval-valued probabilities, possibility measures, plausibility measures, and upper and lower expectations or previsions.

Helton and Oberkampf [?] chaired a workshop in which many of these alternatives were tested against two test problems with very mixed results. Personal discussions with William Oberkampf lead the author to share his conclusions that possibility theory and imprecise probability theory are too undeveloped at the current time for serious VV&A usage. On the other hand, fuzzy simulations have reached a level of sophistication to be applied to engineering problems; see [?]. The Helton and Oberkampf [?] experiment should be pursued.
A comprehensive review of decision-making literature by Zsambok and Klein describes fifteen standard models of decision-making approaches. But those approaches are developed on either Bayesian (subjective) or frequentist accounts. Clearly, extensions to alternative models for decision-making are required.

6. Impact on Modeling and Simulation

Current practice in M&S is to develop a simulation based on a model and then use statistics to validate it. The research proposed here is based on alternative models of arithmetic and the use of uncertain systems [?]. By addressing uncertainty from the beginning of the project, the question of validation is naturally developed within the simulation framework. For a complete example, see Chapter Seven of [?].

The principles are in place for such a program: interval arithmetic is supported in both Fortran and C++ by Sun Microsystems. Correct interval arithmetic directly supports fuzzy arithmetic, which can be used to model probability distributions. In addition, there are paradigms for simulations that can include alternative arithmetic or standard arithmetic, such as simulated annealing and discrete event simulations. Although these technologies are well-developed, they are not in the main stream M&S applications, partially due to the lack of exposure to simulationists and partially due to apparent performance issues.

A simple example of these concepts can be illustrated, in the spirit of Helton and Oberkampf, using a sophomore physics problem:

Consider a compact object falling from the top a cliff for which only approximate height data is known. How long will it take for the object to reach the ground?

While the appropriate physics is known,

\[
\frac{d^2y}{dt^2} \approx -g + D\left(\frac{dy}{dt}\right)^2.
\]

There is a range of values for both \(g\) and \(D\); in fact, it is not even be certain that \(D \neq 0\). Compounding the problem is the possible fuzziness of the height data. Assume \(D = 0\), \(g = [9.7, 9.9]\) meters per second per second and the height as \(h = [119.5, 125.0]\) meters. Using Sun’s f95 interval implementation, \(t = [4.90, 5.08]\). The interval value for \(t\) contains the shortest (4.90) and longest (5.08) times consistent with the input, but it does not indicate the most likely value.
To compute the most likely value, we use fuzzy arithmetic. For this example, triangular numbers are used. Let \( g = \langle 9.7, 9.8, 9.9 \rangle \) with 9.7 being the low bound, 9.8 being the most likely value, and 9.9 being the high bound. Similarly, let \( h = \langle 119.0, 120.0, 125.0 \rangle \). Using this value, \( t = \langle 4.90, 4.95, 5.08 \rangle \). Notice that the fuzzy value is within the interval arithmetic value, as it should be.

If \( D \neq 0 \), then the problem is much more difficult because Equation 1 is solvable for \( t \), but the result is not very intuitive:

\[
s = \frac{\log(\cosh(t\sqrt{bg} - a\sqrt{Dg}))}{D}
\]

For illustration, the problem is solved instead by using a numerical algorithm. We obtain the fuzzy numbers:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Height</th>
<th>Drag</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>119</td>
<td>9.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Confident</td>
<td>120</td>
<td>9.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Maximum</td>
<td>125</td>
<td>9.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Notice that there is a 1.5% error in the minimum time between the algebraic answer and the numerical answer. It turns out that the simple expedient of changing the rounding mode does not solve the problem, indicating that different numerical algorithms are needed. This numerical error also illustrates how uncertainty is naturally introduced into the VV&A process: How does the numerical answer relate to the algebraic answer? What are the risks of using one or the other?

7. Conclusions

Foundations 2004 was organized to report on research in risk, uncertainty, and uncertainty quantification. However, a more fundamental result becomes clear: the understanding that VV&A is ultimately better understood in a decision-making context. Hence, two recommendations are proposed.

**Recommendation 1.** We recommend that VV&A processes be formulated with decision-making research as a guideline. The fundamentals of decision-making are closely related to problem-solving, as made clear by Kahneman, Solvic, and Tversky. However, the research of Zsambok, Klein, and others have shown that the the problem-solving paradigm may not be used in practice
by many decision-makers in certain situations. The psychology of expertise and decision-making continues to be an active area [?]. It is clear that working management with VV&A responsibilities must be able to understand and develop uncertainty models as a natural part of decision-making. Educational programs must be developed to support VV&A practitioners.

We propose that VV&A develop information models based on the standard approaches used in decision-making and based on the problems themselves. For example, CSE models will continue to be evaluated using sophisticated statistical methods developed especially for the problem at hand; we should not expect those methods to work in all situations. There are wealth of tools in the business risk market that bear evaluation.

Recommendation 2. Our second recommendation is that M&S practice focus on modeling and simulation of **uncertain systems**, systems in which uncertainty is a fundamental part of the modeling and simulation design. This focus can be accomplished in several ways.

1. Modeling and simulation education must integrate the uncertainty present in models into simulations that can produce information useful in quantifying that uncertainty. Engineering and science courses, traditional CSE courses, should be modified to include uncertainty issues in models.

2. We must develop trusted software libraries and language platforms that allow simulationists to develop simulations of uncertain systems.

3. Research must be undertaken to explore alternative forms of uncertainty and uncertainty quantification. Klir and Weirman note that the various types of uncertainty quantification are not fully explored. But the various forms of uncertainty quantification can only be used after a model of uncertainty has been developed.

4. The alternative models of uncertainty that have appeared in the literature are not as developed as the axiomatic probability systems. It may well be too soon to know whether or not these alternative systems will provide new insights. One aspect of the Klir and Weirman approach that has been overlooked is the focus on information theories. The author first discussed the decision making paradigm with information focus in [?].

5. We recommend that an organized effort be made to evaluate these alternative formulations to extend the results of [?].

---

1Laplace, “It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.” Théorie Analytique des Probabilités (1812).
We recommend that case studies be developed to educate developers on the cost/benefits of VV&A.

We have implemented an extension to Sun’s f95 compiler that computes fuzzy functions. We can develop supporting libraries, which will require that the appropriate numerical analysis and numerical methods be explored. This development will be driven by the need for tutorial programs.

**Keywords.** Evidence, probability, inverse logic, systems theories.

We describe a new technique leading to quantifiable predictions during fidelity, validation, and verification processes. The technique described uses the structure of the theory under which a model or simulation is developed. We quantify the concept of evidence using odds. Experiments may lead to complex trees from which we can compute odds; the algorithms are based on our definition of inverse logic.

\[
\text{\"A habit of basing convictions upon evidence, and giving to them only that degree or certainty which the evidence warrants \ldots [would] cure most of the problems in the world.\"}
\]

Bertrand Russell.

8. Introduction

We describe work in progress to develop formal methods and tools for verification and validation of models and simulations. Such tools would be useful if they could be integrated into the natural modeling and simulation development cycle and if they could lower costs and enhance quality without being too intrusive. We present a mathematical system based on the idea of *systems theories* and is a natural outgrowth of [?]. From the structure of the theory we can develop statistical measures that quantify the evidence.

Verification has a long history of computer tools. Fidelity, on the other hand, is relatively new issue. Fidelity (called *validation* in the sciences) is defined in [?] as

\[
\text{\"the degree to which a model or simulation reproduces the state and behavior of a real world object or the perception of the real world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation.\"}
\]

In general, fidelity studies will require observations that have inherent uncertainty; simulations may interject uncertainty in many ways. Therefore, we expect reasoning with and about uncertainty to
Validation, as defined in [?], is concerned with both fidelity models and the setting of statistical parameters.

Reasoning with and about uncertainty has a long history. Probability and the scientific method hark from the 17th Century. More recently, artificial intelligence research have produced whole theories of uncertainty (see [?]); *A Mathematical Theory of Evidence* by Shafer [?] and Pearl’s *Causality* [?] deserve special mention. In other formal method approaches, program proofs [?] and model checking [?] are other possible approaches to understanding uncertainty.

Unfortunately, most approaches to dealing with uncertainty require certain probabilities to be available; how to do this is controversial. We focus on one new approach to computing such probabilities using the proof of the theorems as constraints on the probabilities.

We link our investigation to Sargent’s new diagram, Figure ???. Notice that systems theory is the single entity that bridges the gap between the real world and the simulation world, making system theory central to modeling and simulation (M&S) as well as to validation, verification, and accreditation (VV&A). Since systems theories are the central issue, we must relate VV&A terminology to the systems theory paradigm (This is the subject of another paper). System theories are formal mathematical systems, but they can be used for planning and management of the development cycle. The system theory that is assumed by a model or simulation should determine measurable, objective conditions for fidelity and verification activities. Models are often special cases of a more general theory. Simulations can be models themselves or instances of models.

A ready example of a systems theory is linear optimal control theory of dynamical systems used in virtually every aspect of modern technology. The domain of the theory is any object that fits the assumptions of the theory as well as any mathematical axioms laid out. Generally, the objects of the theory fit some form of behavioral representation. The domain is chosen based on a problem to be solved. The output of the theory is a justified, coherent methodology for solving the system problem. This methodology generally concerns either the analysis of behavior or the synthesis of controls (or both). But the theory is formal, relying on properties that may not hold in the real world. Engineers have learned how to adjust for these situations; in a sense, engineers validate by prototype and test, but they are guided by the formal conclusions.

We chose uncertain dynamic systems as the general theory. Uncertain systems [?] were defined in 1973 with continuous development since. Uncertain systems recognize that we must estimate such variables as state and that observations are statistical. These must be linked by a
mathematical model explaining the estimation and the error behavior. Let \( x \) be the true values of the state variable and \( \hat{x} \) be the estimated value. Let \( z \) be the actual observed value and \( v \) be the uncertainty in \( z \). The estimation problem for the theory is

Given the a priori model for the uncertainty \( x, \hat{x}, \) and \( z \), determine a particular a posteriori model for the uncertainty of \( x \) which which involves the “least” uncertainty. [?].

An example of an uncertain system is the discrete time linear system with white noise

\[
x(n\Delta + \Delta) = \Phi(n\Delta)x(n\Delta) + \\
\Delta G(n\Delta)w(n\Delta)
\]

\[
E\{w(n\Delta)\} = 0
\]

\[
E\{x(0)\} = 0
\]

\[
E\{x(0)w'(n\Delta)\} = 0
\]

\[
E\{x(0)x'(0)\} = \Psi
\]

\[
E\{w(n_1\Delta)w'(n_2\Delta)\} = \begin{cases} 
\Delta^{-1}Q(n_1\Delta), & n_1 = n_2 \\
(0), & n_1 \neq n_2 
\end{cases}
\]

where \( w \) is a white noise process. Uncertain systems add a new problem: the state estimation problem. For the above system, the estimation problem can be solved with Kalman filters.

Section 9 develops a method of evidentiary reasoning based on the observed odds. Section ?? illustrates the method by exploring the fairness of a coin. The need to invert implication leads to Section ?? with an example based on the above uncertain system presented in Section ???. The simple approach in this last section must be augmented by more general methods outlined in Section ???. Conclusions and future work are presented in Section ???

9. Probability Theory as Inverse Logic

The goal of modeling is to produce a system theory concerning the objects of the theory. Such a theory is primarily logical in that the theory is justified through proof. These proofs produce the ability to justify various referents: measures of state or behavior predicted by the theory. Fidelity, and by subsumption, validation, seeks to quantify the degree to which the observed systems actually predict the referent values.
Let \( \{H_1, H_2, \ldots, H_m\} \) be the hypotheses of the systems theory and \( \{C_1, \ldots, C_n\} \) be the conclusions. Then the methodology and theorems of the theory link the two through a line of reasoning.

\[
\frac{H_1, H_2, \ldots, H_m}{C_1, \ldots, C_n} \text{ (Methodology, Theorems)}
\]

The fidelity question is the following. If we observe an object, say \( x \), that satisfies the conclusions \( C_j(x) \) what can we say about \( H_i(x) \)? Section ?? describes a formalization of the term \emph{information}. Using that definition we define \emph{evidence} in Section ?? and how to compute the collected evidence as a measure of fidelity. Section ?? provides a elementary example using an elementary probability example: Given a sequence of heads and tails from a coin, how can we judge the fairness of the coin?

9.1. **Information.** We begin with the concept of \emph{information}. The term has several interpretations both formally (such as \emph{information theory}) and informally. We build on the works of [?, ?, ?, ?, ?]. Information is a measure of our knowledge of the state of the system after an event relative to our overall knowledge of the state. It is a measure of relevance.

We first consider the \emph{likelihood} function:

\[
L(H : C|G) = \frac{P(C|H \& G)}{P(C|G)},
\]

where \( L(H : C|G) \) is read “the likelihood of \( H \) in light of event \( C \) given global knowledge \( G \).” Normal use of the term \emph{information} leads to the interpretation that if the likelihood is one, we would say there is no information:

\[
I(H : C|G) = \log_b L(H : C|G).
\]

The base \( b \) of the logarithms is immaterial allowing us to use “natural” units, such as decibels and bits.

9.2. **Evidentiary Reasoning.** Following [?, ?], we want to measure the evidence available based on the information at hand. In our case, this is the difference in the information for a state minus
the information available for not being in that particular state. In other words,

\[ W(H : C|G) = I(H : C|G) - I(\bar{H} : C|G) \]

(4)

\[ = \log_b \frac{P(C|H\&G)}{P(C|\bar{H}\&G)} \]

(5)

\[ = \log_b F(H : C|G), \]

where \( \bar{H} \) is the complementary state of \( H \). Notice that \( W \) is naturally stated as “odds”. Using Bayes theorem in its odds form we get

\[ O(H : C|G) = O(H|G)F(H : C|G) \]

and the log odds as

\[ \log_b O(H : C|G) = \log_b O(H|G) + \log F(H : C|G). \]

We take this to be our computational rule for evidence

\[ e(H : C|G) = e(H|G) + \log F(H : C|G) \]

\[ = e(H|G) + W(H : C|G). \]

In order to make the summation start at zero, we agree that the evidence at the beginning of a study is zero by taking the initial odds to be 1.

10. Example

We choose a elementary probability problem to illustrate evidentiary reasoning: coin tossing of a fair coin. Coin tossing often shows up before either conditional probability or Bayes theorem are studied formally because our intuitions are generally easy to examine.

In our context we define

\[ P(Head|Fair) = \frac{1}{2} \]
\[ P(Tail|Fair) = \frac{1}{2} \]
\[ P(Head|Unfair) = \frac{1 - 2\epsilon}{2} \]
\[ P(Tail|Unfair) = \frac{1 + 2\epsilon}{2} \]
Mathematics has taken the “easy” direction: The experimenter takes a coin known to be fair and runs experiments. The validation question is harder: “Given a sequence of coin tosses, how can I decide whether or not the coin is fair?”

10.1. Development. In this example, it is easy to compute the factor directly

\[
F(Fair : Head|G) = \frac{P(Head : Fair|G)}{P(Head : Unfair|G)} = \frac{1}{2} \cdot \frac{1 - 2\epsilon}{2} = \frac{1}{1 - 2\epsilon}
\]

On the other hand,

\[
F(Fair : Tail|G) = \frac{1}{1 + 2\epsilon}
\]

\(\epsilon\) can be between \([0, 0.5]\). If \(\epsilon\) is zero, then the coin is fair and the sum of the evidence is zero. But if \(\epsilon\) is 0.5, then any head is impossible and \(F(Fair : Head|G) = \infty\) while \(F(Fair : Tail|G) = 1/2\).

The weight of evidence is logarithm of the factor, so

\[
W(Fair : Head|G) = -\log(1 - 2\epsilon)
\]

\[
W(Fair : Tail|G) = -\log(1 + 2\epsilon)
\]

If we use common logarithms, we are measuring evidence in bels; multiply by 10 and we get decibels, a measurement commonly used in engineering. The author generated 100 samples of 20 random numbers each from the binomial distribution with \(p = 1/2\). This is the \(\epsilon = 0\) case. The total evidence is the sum which turned out to be 30.58 (bels) for this sample set. That is, is 306 decibels a good number or a bad number? 306 is one chance in \(10^{30}\) that the coin is not fair.

A second experiment using the same rubric was used. In this second example we intentionally choose a distribution that could not be a binomial: the log normal with mean 0.45 and variance 0.14. This is decidedly spiked around 0.45. In this case, the evidence was -618. Again, the evidence is off the chart, but as evidence against.

Result. We have shown that if we were to observe the first 2,000 trials we would most likely conclude that our theory is valid. The second set of 2,000 trials we would be forced to conclude that our theory was wrong or that we did not have a fair coin. Notice that we cannot conclude that the coin is unfair. Why is this?
11. Inverse Reasoning

Our conclusion comes from the laws of logic. While it is true that we can say “If \( a \rightarrow b \) and \( a \), then conclude \( b \)” it is also true that “If \( \neg a \), then \( b \). Since we are reasoning “backward” from the conclusions, we must deal with both situations.

George Pólya considered many of these concepts in [?]. To understand Pólya, we need to understand that Bayes Rule is an application of standard implication.

If \( A \rightarrow B \) then

\[
P(B|A)P(A) = P(A|B)P(B)
\]

Since \( B \) is a consequence, the left conditional is 1.

\[
P(A) = P(A|B)P(B)
\]

Solving for \( P(A|B) \):

\[
P(A|B) = \frac{P(A)}{P(B)} \quad (6)
\]

\[
O(A|B) = \frac{P(B)}{P(B)} \quad (7)
\]

\[
(8)
\]

In other words, if \( A \rightarrow B \) and we observe \( B \), then \( A \) is more plausible by the ratio of \( P(A)/P(B) \).

In odds terms, only \( B \) shows up.

12. Evidence Calculations in Static Uncertain System

A static model of uncertain dynamic systems is described by equations

\[
z = Hx + v \quad (9)
\]

\[
E\{x\} = 0 \quad (10)
\]

\[
E\{v\} = 0 \quad (11)
\]

\[
E\{xx'\} = \Psi \quad (12)
\]

\[
E\{vv'\} = R \quad (13)
\]

\[
E\{xv'\} = 0 \quad (14)
\]
This represents the observation of a system that is static but the observation is uncertain. Using elementary probability considerations, we can show:

\[ E(z) = 0 \quad \Gamma_z = H\psi H' + R \]

This presentation makes clear the relationships among the assumptions and the conclusions. The model also makes clear that there are only two possible observations: \( E(x) \) and \( \Gamma_z \). Suppose now that these are measured with\( \mathcal{O}(E(z) = 0) = d_1 \) and \( \mathcal{O}(\Gamma_z = H\psi H' + R) = d_2 \). What can we conclude about the odds of \( E(z) \) being true?

Using \( E(z) \) in odds form, we see that we have

\[ d_2 = \mathcal{O}(E(z) = 0 & E(z) & E(z) & E(z) & E(z)) \]

\[ d_1 = \mathcal{O}(E(z) & E(z)) \]

If we adopt a “no information” principle then

\[ \mathcal{O}(E(z) | E(z) = 0) = \sqrt{d_1}. \]

For such a simple structure, this informal approach is adequate. What about the more complex cases likely to arise in practice?

13. General Solution

In the general case, the information presented by the systems theory and any proofs in that theory are likely to be very complex. It is generally true that proofs will define a tree \([\ast]\) that can be used to generate information. This is a common approach in artificial intelligence. Perhaps the most complete version of the artificial intelligence approach is in Pearl \([\ast]\). However, the general approach traverses the tree in a forward direction — in the verification direction.

It turns out that George Boole devoted almost one-third of his famous book \([\ast]\) to probability. Boole’s goal was to develop methods by which any set of equations with any given set of prior probabilities could be solved for the remaining probabilities. Boole’s methods were highly criticized during his lifetime; it would be more than a century for his method to be vindicated \([\ast]\). The thread of his ideas is now known as probability logics, among other names. Genesereth and Nilsson \([\ast]\, Chapter 8\) describe a general solution technique in Sections 8.3 – 8.7. Briefly, the general solution
is to develop an optimization problem which describes the constraints on the probabilities of the
system. We can then find an optimal solution. Since probabilities and odds are inter convertible
by
\[ p(Q) = \frac{O(Q)}{O(Q) + 1}, \]
we can use the same technique to solve for the odds of any event. From this, we can determine the
evidence needed for the validation problem at hand.

This technique offers an important advantage: it is adaptive. The basic process is:

1. Define the graph of the validation problem to be the proof tree described above.
2. Using the rules laid out by Boole, we can convert the graph to an optimization problem in
   odds.
3. Solve this optimization problem for the odds given the observations. The optimization
   problem can be augmented by cost and risk data.
4. Solve the evidence problem for this initial graph.
5. Iterate.

14. CONCLUSION — THE BASIS FOR FORMAL METHODS

We have developed an approach to validation that is both logically sound and leads to
quantifiable predictions. The graph of the application can be processed in either direction. The
basic process is

(1) The system theory under which the model or simulation is identified.
(2) Questions concerning the model or simulation are developed by considering the proofs in
   the theory. Proofs can always be represented by a tree.
(3) Questions concerning verification are processed in the from the assumptions and axioms
   toward the conclusions.
(4) Questions concerning validation are processed in reverse using Boole-Pólya-style rules. This
   results in an optimization problems solvable by standard methods.

Work remains to be done in the complete specification of the inverse logic. For example,
Bayes conditioning is not exactly the same as implication. From logic, the definition of implication
is \((A\&\neg B)\) which has a probabilistic rendering of
\[
P(A \rightarrow B) = 1 - P(A)P(\neg B)
\]
\[
= P(\neg A) + P(B) - P(\neg A\&B).
\]

Work dealing with boolean equations [?] may contain algorithms to effectively compute such expressions.

Coherence Consistency (Topological)

Credibility Justifibility (Proof/Demo)

Lawful Nature Methodological Foundation

Organized Nature Reasoned Basis (Logical)

Relevance

\textbf{Part 1. Pólya Rules}

Let’s consider an expansion of Pólya’s rules for plausibility.

\textbf{15. General}

Whatever is in logic should be in the probability rules. This includes

(1) The boolean operators

(2) The rules of inference

(3) The introduction and discharging hypotheses.

The truth tables used to illustrate the rules need to be organized first by independent names such as \(A\), then by functions. The independents and the others should be set off by a double vertical bar to clearly demark the independent and dependent uses.

There is a need to specially treat subjective implication and to interpret the material implication tables in the subjective mode. Example: for the TFTT column of implication, the interpretation is: top two rows are “modus ponens, spurious” with row 3 equal “inconsistent” and row 4 as spurious.

Last but not least, there is a difference between the operator and its application. This is especially true of implication: \(P(B|A)\) deals with the implication but \(P(B|A)P(A)\) is the application of \textit{modus ponens}.
Real World & Simulation World Relationships in Developing System Theories and Simulation Models with Verification and Validation (V&V) -- additional explanation in annotation

Notes:
- Experiment objectives should derive from validated requirements
- Dotted red implies comparison, assessment, or evaluation
- Validation is always relative to objectives, requirements, and intended use

Diagram by Robert G. Sargent (Syracuse U) Jan 01
This is the simplest case examined above. If \( A \rightarrow B \) then we have

\[
P(B|A)P(A) = P(A|B)P(B)
\]

Since \( B \) is a consequence, the left conditional is 1.

\[
P(A) = P(A|B)P(B)
\]

Solving for \( P(A|B) \):

\[
(15) \quad P(A|B) = \frac{P(A)}{P(B)}
\]

Here we have to issue a warning about the assumption If \( A \rightarrow B \) then \( P(B|A) \) is 1. The problem is that if we use *material implication* as used in mathematics and mathematical logic, then it is perhaps meaningless. If we use *subjective implication* as used in science and engineering, then the statement is correct. \( P(A|B) \) is, in some sense, the confidence in the proposition \( A \) is more creditable now that \( B \) has been “confirmed.”

**Need some examples**

Consider now the total differential of the Eq ???. From elementary calculus,

\[
(16) \quad \partial P(A|B) = \frac{\partial P(A)}{\partial A} \Delta A - \frac{\partial P(B)}{\partial B} \times (P(B))^{-2} \Delta B,
\]

where \( \Delta A \) and \( \Delta B \) are the changes in plausibility. Therefore, if \( P(A|B) \) is constant, \( P(A) \) and \( P(B) \) change in the same direction, meaning the increase in plausibility of \( B \) increases the plausibility of \( A \).

Let \( P(B) \) be constant. From Eq ?? we can see that \( P(A) \) cannot be greater than \( P(B) \). This guarantees that our faith in \( A \) cannot be greater than \( B \), thus eliminating over-exuberance. From the other standpoint, \( P(B) \) cannot be less than the probability of \( A \) since \( A \rightarrow B \).

Lastly, Eq ?? can be solved slightly differently. Let

\[
P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})/
\]
Situation Assumptions Formula
Consequence I \( A \rightarrow B \) \( P(A|B) = P(B)P(A|B) \)
Consequence II \( P(A|B) = P(A)/P(B) \)
Consequence III \( P(A|B) = P(A)\left(1 + (1 - P(A))P(B|\bar{A})\right) \)
Grounds \( A \) is implied by \( B \) \( P(A|\bar{B}) = \frac{P(A) - P(B)}{1 - P(B)} \)
Incompatability \( P(AB) = 0 \) \( P(A|\bar{B}) = P(A)/(1 - P(B)) \)
Sequential See Evidence
Circumstantial Evidence \( P(A|B) > P(A|\bar{B}) \) \( P(A)/P(A|B) = P(A) + (1 - P(A))P(B|\bar{A})/P(B|A) \)

Figure 2. Table of Pólya Results

Then Eq ?? becomes

\[
P(A|B) = \frac{P(A)}{P(A) + P(A)P(B|\bar{A})}.
\]

Start with \( P(A) \) constant. The variable portion is \( P(B|\bar{A}) \).

17. EXAMINING A POSSIBLE GROUND

**Problem Set Up.** We are interested in conjecture \( A \). We notice that \( B \) is a possible ground for \( A, B \rightarrow A \). Suppose that \( B \) is not substantiated by experiment. How should our confidence in \( A \) be changed?

This can be derived as above and we get

\[
P(A|\bar{B}) = \frac{P(A) - P(B)}{P(B)} = \frac{P(A) - P(B)}{1 - P(B)}.
\]

The lefthand side is the credibility after \( B \) is refuted and the righthand side is the situation before \( B \) is refuted. **This is the non-odds version of evidence.** Explore with respect to \( P(A) \) as a constant and \( P(B) \) as a constant; vice versa;

18. EXAMINING A CONFLICTING CONJECTURE

Conflicting means \( P(AB) = 0 \). Work it out and explore.

19. EXAMINING SEQUENTIAL CONSEQUENCES

This should be seen as the evidentiary formulas.

20. CIRCUMSTANTIAL EVIDENCE

The expression is “\( B \) with \( A \) is more credible than without \( A \). Model this as \( P(B|A) > P(B|\bar{A}) \).
<table>
<thead>
<tr>
<th>Form</th>
<th>Input Values</th>
<th>Alternative Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a = 0 \ 0 \ 1 \ 1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b = 0 \ 1 \ 0 \ 1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>false</td>
<td>0 0 0 0</td>
<td>false, zero, clear</td>
</tr>
<tr>
<td>( a \land b )</td>
<td>0 0 0 1</td>
<td>and</td>
</tr>
<tr>
<td>( a \land \neg b )</td>
<td>0 0 1 0</td>
<td>and-not, inhibit, ( a &gt; b )</td>
</tr>
<tr>
<td>( a )</td>
<td>0 0 1 1</td>
<td>( a )</td>
</tr>
<tr>
<td>( b \land \neg a )</td>
<td>0 1 0 0</td>
<td>inhibit, ( a &lt; b )</td>
</tr>
<tr>
<td>( b )</td>
<td>0 1 0 1</td>
<td>( b )</td>
</tr>
<tr>
<td>( a \oplus b )</td>
<td>0 1 1 0</td>
<td>xor</td>
</tr>
<tr>
<td>( a \lor b )</td>
<td>0 1 1 1</td>
<td>inclusive or</td>
</tr>
<tr>
<td>( a \lornot b )</td>
<td>1 0 0 0</td>
<td>nor</td>
</tr>
<tr>
<td>( a \oplus \neg b )</td>
<td>1 0 0 1</td>
<td>if and only if, equality, equivalence, xnor</td>
</tr>
<tr>
<td>( \neg b )</td>
<td>1 0 1 0</td>
<td>( \neg b )</td>
</tr>
<tr>
<td>( a \lor (\neg b) )</td>
<td>1 0 1 1</td>
<td>or-not, implication, ( b \Rightarrow a )</td>
</tr>
<tr>
<td>( \neg a )</td>
<td>1 1 0 0</td>
<td>( \neg a )</td>
</tr>
<tr>
<td>( b \lor (\neg a) )</td>
<td>1 1 0 1</td>
<td>implies, implication, ( a \Rightarrow b )</td>
</tr>
<tr>
<td>( a \lornot b )</td>
<td>1 1 1 0</td>
<td>nand</td>
</tr>
<tr>
<td>true</td>
<td>1 1 1 1</td>
<td>true, one, set</td>
</tr>
</tbody>
</table>

FIGURE 3. The Standard Boolean Functions
APPENDIX B. REASONING RULES

Tautological Implications

\[ \frac{A \rightarrow B \quad A}{B} \] MPP

\[ \frac{A \lor B \quad \neg A}{B} \] MTP

\[ \frac{A \quad B}{A \land B} \] Adjunction

\[ \frac{(A \land B) \rightarrow C \quad A \rightarrow (B \rightarrow C)}{A \rightarrow (B \rightarrow C)} \] Syllogism

\[ \frac{A \rightarrow (B \rightarrow \neg B) \quad \neg A}{\neg A} \] Absurdity

\[ \frac{A}{A \lor B} \] Addition

Tautological Equivalences

\[ \frac{A}{\neg \neg A} \] Double Negative

\[ \frac{\neg A}{A} \] Double Negative

\[ \frac{A \rightarrow B}{\neg B \rightarrow \neg A} \] Contrapositive

\[ \frac{\neg B \rightarrow \neg A}{A \rightarrow B} \] Contrapositive

\[ \frac{\neg (A \land B)}{\neg A \lor \neg B} \] deMorgan I

\[ \frac{\neg A \lor \neg B}{\neg (A \land B)} \] deMorgan I

\[ \frac{\neg (A \lor B)}{\neg A \land \neg B} \] deMorgan II

\[ \frac{\neg A \land \neg B}{\neg (A \lor B)} \] deMorgan II

\[ \frac{A \land B}{B \land A} \] Commutation

\[ \frac{B \land A}{A \land B} \] Commutation

\[ \frac{A \lor B}{B \lor A} \] Commutation

\[ \frac{B \lor A}{A \lor B} \] Commutation

\[ \frac{A \rightarrow B}{\neg A \lor B} \] Definition of ‘→’

\[ \frac{\neg A \lor B}{A \rightarrow B} \] Definition of ‘→’

\[ \frac{(A \rightarrow B)}{A \land \neg B} \] Negative Implication

\[ \frac{A \land \neg B}{(A \rightarrow B)} \] Negative Implication
We provide an introduction to the concepts of validation, the process of determining whether or not a model or simulation adequately represents natural processes. We describe a formal logical mechanism due to Carnap and Hempel which has been used to study science models. Finally, we apply our concepts to the conduct of the Michelson-Morley experiment. This experiment was supposed to validate the concept of æther; it failed, opening the way for relativity.

Keywords. Validation, physics, logic, numerical analysis.

APPENDIX C. MOTIVATION

For most of recorded history, the tools of mathematical analysis have been sufficient to investigate the behavior of models. This changed in the 1960’s when Edward Lorenz stumbled, literally and figuratively, into chaos. Such systems behave in very nonintuitive ways, making the task of validation very difficult. These studies are now carried out on computers using simulations.

Discrete event simulations suffer from the same growth in complexity. Consider any computer communications system. Demands for speed and reliability make every model expensive due to detail and data gathering. Validations are even more difficult when people are “in the loop”.

But computer simulations are not the only problem. The recent Cerro Grande forest fire in the Bandelier National Forest near Los Alamos, New Mexico, in May, 2000 is an example. Los Alamos National Laboratory had had a forest fire in 1996. The Laboratory and the U. S. National Park Service developed a simulation to predict the safety of “prescribed burns.” In the Cerro Grande fire, the inquiry board found that the inputs as specified were quite correct, but the model itself was poorly conceived. Furthermore, the human organizations did not collaborate in ways that would have prevented the fire.

Let’s agree to call simulation any use of computation to solve a behavior posed by a model. For much of its history, simulation has been a second choice for developing and testing products and services. In the 1990s, business competition made mistakes.
prohibitively costly in both time and resources. The United States Department of Defense is seriously considering a procurement policy based on computer simulations, not prototypes. Science, too, is using simulations in a more pivotal way. Experiments in science have become ever more dangerous or completely impossible to carry out: e.g., nuclear winter and the maintenance of the U.S. nuclear weapons stockpile.

Whether discrete or continuous, modeling our world has more difficult due, in large measure, to what Perrow calls “mind-boggling complexity.” This complexity manifests itself in simulations by massive amounts of code which is impossible to document or maintain. We must insure that any decision based on the output of a simulation is grounded in reality.

Our purpose is to introduce the reader to two processes: validation and warranty warrants for credibility, capability, and relevance. Due to the inductive nature of science, complete validation is impossible. A more practical approach to validation is to establish the correctness of specific uses of a model or simulation with restricted parameters; this second approach I shall refer to as warranty. In this view, validation is an infinite number of warrants.

This paper has three parts. Part I discusses the validation and warranty processes as described in the literature. Part II discusses logical and mathematical issues relating to the conduct modeling and simulation. Part III applies the developed concepts to the famous Michelson-Morley experiment.

Part 2. The Practice of Validation and Warranty

APPENDIX D. PRELIMINARIES

Ultimately, the concern is for the correctness of models. Models are how we come to know our world. Simulations are defined within our understanding of the model.

D.1. Modeling is Key. Modeling comes from our need to understand our world. How deeply we need to understand — hence, how intricate the model — is a function of the costs and risks posed by the problem to be investigated. The costs and risks

---

2 of guaranteeing a model or simulation for a particular use. In the U.S. Department of Defense, the terms accreditation and certification are often used but these have a different connotation.
in turn generate a need to make decisions. Therefore, modeling starts with the real world and a problem and ends in decisions with real world ramifications. Modeling is the process of studying a system, then writing down descriptions of the objects and their relationships. This is a fundamental cognitive process. We endorse John Kemeny’s view that “Science must start with facts and end with facts, no matter what theoretical structures it builds in between.” Modelers must decide what is important to the problem and ignore other aspects. As Richard Bellman said, “…the scientist, like the pilgrim, must wend a straight and narrow path between the pitfalls of oversimplification and the morass of oversimplification.”

D.2. Model Validation and Warranty. Zeigler in [?] remarks, “The essence of modeling is in establishing relations between pairs of systems.” Validation is establishing the truth of those relations between the theoretical and the observed systems.

The term validation means a process of determining the degree to which our model predicts reality. On a purist note, a model can only be validated on real phenomena data although one often hears the term applied to theoretical models (such as engineering design models) that are not tested against reality. This confusion arises because the terms validation and verification are used interchangeably in informal discourse. Several terms are used informally:

1. Calibration is putting numbers on dials.
2. Verification is justifying that one has followed the rules correctly: “did we build the system right?” Verification is an act of justification, like proof.
3. Validation is testing our models against the real phenomena: “did we build the right system?” Validation is about credibility.
4. Warranty is testing the capability and relevance in a context narrower than total validation: “Are we assured that the system is right for this problem with these particular parameter settings?”

A full validation would be the set of all relevant warrants. Therefore, the overall scheme of a validation plan is to (1) establishing the warrants for credibility, capability, and relevance; (2) establishing a plan to collect the data specified; and (3) comparing the values produced to the observations.
D.3. Basics of Validation. Reality includes intended use, costs, risks, and benefits. Validation would be trivial if we lived in a perfect, Platonic world. In such a world

1. All observations were perfectly taken without any error or imprecision;
2. All parameters were perfectly known;
3. All physical relationships were perfectly known and functional;
4. All calculations could be done at infinite precision.

Alas, we do not live in such a world. The nature, depth, and breadth of information are outside the application’s control. Therefore, we must establish that the model as proposed (and perhaps implemented in a simulation) has the credibility, capability, and relevance to allow us to make correct decisions.

D.4. Simulation Implements Models and Is Not Modeling Itself. Most useful models are eventually too complicated to be explored by hand. Even assuming that one has a model that completely captures the system under study, it is by no means assured that one can write a computer simulation for all possible situations described by the model. One issue is the finite capabilities of the computer, which leads to problems of imprecision. The larger problem is the whole development of the software [?, ?]. There are some techniques, as we discuss in the Best Practices, Section ??.

While the technical details of modeling and simulation are quite different, they are attempting to do the same thing. The author has heard of some astrophysics codes for which the code is the model. We focus on the common ideas within modeling and simulation.

There is some tendency for workers to equate uncertainty with numerical error. This may be because of the emphasis on numerical error in numerical analysis and numerical methods texts. Suppes comments in [?] “It is common for models of a theory to have continuous functions and infinite sequences although confirming data is highly discrete and finitistic in character.” This is even truer for computers. The importance of this remark lies in the great loss of information going from the infinite to the finite. Uncertainty is the lack of information. In model validation, the study of uncertainty is a central issue. But uncertainty about what?

1. Observations: precision of data, completeness, erroneous data;
(2) Predictions: functional form;
(3) Decisions: are all factors accounted for? Are all criteria accounted for? Science and engineering add another problem: phenomena usually have multiple explanations (theories).

D.5. Conditions for Validation. One way to think about validation is to consider the subjective syllogism. Observation produces the concrete data after the appropriate statistical procedures have been applied. Predictions come from the simulation. The predictions then must be compared to real data. Because of uncertainty, we must make the earlier cited Zeigler decision with uncertainty. Research in VV&A frameworks must properly deal with the two different reasoning styles in science and mathematics: (1) inductive reasoning and (2) deductive reasoning. To fully grasp the confusion and controversy, one must first understand the differences in scientific and mathematical truth.

Science uses inductive reasoning (not the same induction as used in high school induction proofs). In inductive theories, we have little or no formal logic help. In effect, scientific theories are proof by affirming the consequent which is a fallacy in deductive logic. Inductive systems are such that the can find new information from the premises. On the other hand, deductive logic, so familiar to computer scientists, mathematicians and engineers, cannot find new information. In fact, the premises must hold all the keys to proving the consequences.

Thus, to understand validation, we must look to the scientists view.

D.6. Understanding the Scientists’ View. The major distinction between science and mathematics is that the scientists look for explanations that can be related to the “common man.” Mathematics is used as a tool, but the equations have intrinsic meaning whereas to a mathematician, an equation is a set of symbols with no intrinsic meaning. Here, in Feynman’s words, is the crux of the matter. “But, for example, in Faraday’s law we don’t really need mathematics. In this case, Faraday’s law is ‘obvious’ manifestations of atoms. But for Newton’s law of gravitation, we only have mathematics because the underlying physical causes are not known in detail. Mathematics is needed for detailed phenomena in large, complex systems” [?].
Knowledge is our ability to explain. But that explanation must be accepted by the scientific community as a whole. Despite the popular perception of the lone researcher, the methodology of science is an agreed upon set of principles used to build a coherent set of explanations. Knowledge requires that we justify our conclusions using the methodology. Unfortunately, the bosses’ view of what makes scientific work acceptable and the communities’ view of acceptance are vastly different [?].

D.7. Explanations and the Multiple Assumptions of Science. Open any science and engineering textbook. A quick reading will convince the reader that there are multiple sets of competing assumptions about any subject. We call these sets of assumptions explanations. In theory, there is a language system for each such explanation.

On the other hand, the need for the scientist is to consider competing explanations as a system. That is, we must be able to reason about the various views resulting from a particular set of observations.

Scientists and engineers look for mechanism. It is not enough to have an equation describing something unless the equation is understood to be meaningful. This meaning lies in the consistency of explaining phenomena.

D.8. Real Explanations. Explanations are not justifications. “Explanation is a basic feature of science. Science seeks to organize and systematize our knowledge of what goes on in the world on the basis of explanatory principles that can afford answers to why questions” [?]. Explanations must elucidate linkages between the explanans (the thing doing the explanation) and the explanandum (the thing being explained). It should elucidate how the connections among objects actually function. They should be at least coherent with current understanding; that is, they should integrate laws into a corpus, an accepted scheme of things.

There is a constant problem in science that deals with causality [?]. In general, scientists want to deal with causality, while mathematics and logic in general, and statistics in particular, do not. This will be a continuing problem.

Explanations must make the subject intelligible, usually by reducing it to the familiar. The understandability of the explanation is the key concept. In the scheme of things, it may be necessary to add concepts of process: the how of the subject.
Explanations should provide the principal considerations bearing on correctness and relative strength of the explanation. The explanation should be complete — it should explain all observations. Explanations must clarify predictions.

The explanations must subsume observed events within accepted, established generalizations and laws; i.e., the scientific methods of the discipline. Here we get into the probabilistic explanations versus universal law. “A scientific explanation will be a subsumption argument whose major premise is a suitably lawful proposition” [?, p. 13]. Consistency shows here: Subsumption — a synonym for “is a subset of” — demands topological considerations. Regardless of the form of the explanation, Feynman sets the goal of validation: “Every theory that you make up has to be analyzed against all possible consequences.” [?, p. 39].

Therefore, validation of model is relative to the underlying explanation it implements.

D.9. Scientific Reasoning. Mathematical thinking can actually be a hindrance because equivalent formulations in physics add insight; mathematical equivalence (isomorphisms) does not. “Mathematics is a language plus reasoning — Mathematics is a tool for reasoning. It is in fact a big collection of the results of some person’s careful thought and reasoning. But with mathematics it is possible to connect one statement to another” [?, p. 40]. Physical laws are not just statements. “Laws of theory extend beyond the range of their deductions.” [?, p. 50]

But the big difference is that in physics, equations have true semantic meaning — their physical manifestation. “We can deduce often from one part of physics, like the Law of Gravitation, a principle which turns out to be much more valid than the derivation (angular momentum). ... We have these wide principles which sweep across the different laws, and if we take the derivation too seriously, and feel that one is only valid because another is valid, then we cannot understand the interconnections of the different branches of physics” [?, p. 49]

Summary. We have examined the modeling process with an eye towards understanding the difference between the theoretical model and the observed system. These differences lead to an uncertainty about whether or not the model has any meaning
that is, is the model a consistent predictor of the phenomena observed. At some point, we can no longer “just talk about it.”

We turn to the question of practice.

APPENDIX E. BEST PRACTICES

We now turn to the problem of what the current practice in validation. The U. S. Department of Defense (DOD) is perhaps the world’s single largest consumer of models and simulations. Because of the unique status of DOD and its unique services, there is a requirement for uniformity of approach to modeling and simulation. The Defense Modeling and Simulation Office (DMSO) is a clearinghouse for information. DMSO has develop documents on best practices [?] that are available on-line through their website: www.dmso.mil. This section explores [?].

Here is a synopsis of the issues.

(1) A problem has been formulated. In the analysis of the problem the elements of possible decisions are enunciated.

(2) The decision makers must choose a solution to the problem by choosing from among the possible decisions. There is uncertainty about what course to choose.

(3) This generates the need for quantitative and qualitative information. There are no certain answers. A series of questions is formulated. These questions are the specification for the models. The information must be known to within a certain accuracy and the values are sensitive to various criteria.

(4) The simulations are written with the model in ?? as the specification. The outputs of the model are used to compute the values need to make decisions.

Validation seeks to answer the question, “How much faith do I put in the generated information?” How do workers now operate on the validation question? DMSO describes three criteria : capability, credibility, and relevance.

In the real world, there are not enough resources to do a complete validation study. Warranty requirements detail the type and depth of information that is needed to ensure that a model meets the modeling requirements; i.e., can present a credible defense of the model. We need to validate observational data as well as simulation data/operation.
To make any validation feasible, the problem must be restricted in realistic ways. Under these restrictions, we talk about warranty rather than validation. Warranty then looks at five issues: (1) functional requirements, (2) fidelity requirements, (3) operating requirements, (4) risks, and (5) budgets.

One role for DMSO is the documentation of practices within the U. S. defense establishment and its customers. To complete our summary we look at the DMSO Best Practices list [?]. Of the techniques listed, the three columns named Informal, Static, and Formal have to be disregarded because the do not include real data input. We have discussed how the formal column should be disregarded since all the elements listed are deductive and not inductive. This leaves of the section of the table listed as Dynamic as usable.

Appendix F. Validation Studies

Generating a validation study is a difficult task. There are, usually, both a model and a simulation. Each has been developed over a period of time. If there is a simulation involved, then this task becomes even more difficult because of the interdisciplinary nature of the product and the team. Generically, these teams are constituted from the application organizations, the algorithm developers (both mathematical sciences and computer sciences) and the architecture groups such as hardware, compiler, software developers and so on. What guidance can we give this team?

The application group would be expected to develop the critical parameters of the application, but the algorithm and the architecture may interact in a meaningful way. For example, in [?] I used a real problem to motivate verifying the arithmetic parameters and their suitability for the task at hand. Another issue that is primarily the purview of the application is the set of critical scenarios that must be simulated before any validation or warranty decisions can be made. From these scenarios, the team would develop a set of metrics for the decisions to be made. The scenarios also can be tested for both costs and risks. Such evaluation should order the scenario execution.

\textsuperscript{3}DMSO has recently updated this document.
The encoding of the decision process is the most crucial since this is where the real observations and the computed predictions meet. The critical development would set the thresholds for the decisions. It is here that the sensitivity analysis plays a role. Sensitivity analysis would identify those values that most effect the outcomes. One must be careful not to over-focus on the model and ignore operating (machine) and physical environmental issues.

The development of the decision criteria will define relationships and from these relationships the metrics that are crucial to the decision. These metrics are often called the *metrics of merit*. These measures can be categorized into two categories: primary measures and dependent measures. The primary measures are used to compute the dependent measures. The outputs can then be categorized in the same manner. The costs and risks go with the primary outputs. For example, the age-old observation about nuclear winters: it is very costly to annihilate the Earth to validate a nuclear winter simulation.

At the end of the design process, the most important question to be answered is, “Do we have consensual agreement that the process here is acceptable to the community?” The answer to this question revolves around agreement on three issues: (1) functional requirements, (2) fidelity requirements, and (3) operational requirements. Only fidelity is discussed here.

Fidelity requirements define the degree of correlation between model and reality. There are two separate issues: (1) accuracy and precision and (2) sensitivity analysis. Accuracy and precision are mostly verification issues. Sensitivity analysis is important both for planning and for the final decision. The basic design issues are the budgeting of computational error onto the measures of merit. The budgeting process will indicate how sensitive the system can be.

The term sensitivity analysis is not uniquely defined mathematically, but revolves around investigations of the linear term of the behavior function. How this is conventionally carried out varies from discipline to discipline.

F.1. The Interaction of Decisions and Verification. The development of the scenarios also put demands on the verification process by specifying which functional capabilities must work. Logical verification is the massaging of the assumptions, limitations, known
errors, and approximation to determine if the model can reasonably be expected to produce results realistic enough to satisfy the information needs. The functional testing can be used to run some validation values or to set the calibration as well as to test.

Verification and validation/warranty are thus intertwined about

1. The correlated decision analysis and functional tests;
2. Audits: overflow and underflows; array bounds;
3. Verified outputs versus hand calculation or verified alternative; and
4. Testing output as model predictions.

F.2. Putting It All Together. The eventual test plan must describe a theoretical test that could be used to measure the actual parameters required to validate a particular functional entity. To this end, it must address the following issues:

1. Calibration,
2. Sensitivity analysis,
3. Topological analysis,
4. Statistical analysis and preparation of observations and simulation output,
5. Statistical Decisions
6. Analytical significance: the model has to be analytically useful. That is, it must actually be good for the purpose.

APPENDIX G. PRELIMINARY CONCLUSION

Validation and warranty are completely separate concepts from verification. In verification we attempt to justify that our simulation fits the requirements implied by the model. Validation and warranty seek to determine how well the model itself fits reality. The three processes are intertwined in practice, so it is imperative that there be close coordination between the developers and the testers. Simulations are moving targets since their use in the decision making process of science, engineering, and management is moving. The use of an accredited model today may be inappropriate tomorrow. In the past, the day to day situation has not been changing as quickly as today’s. The rapidly changing conditions force us to rely on outputs of simulations, perhaps as the sole predictive tool. Therefore, validation and warranty methodology
should be seen as a critical part of the model; indeed, it must become a critical part of the education of the modelers themselves.

We now turn to some technical issues: the analysis of logical relationships and measures of computational goodness.

Part 3. Principles from Logic and Mathematics

To understand the ins and outs of validation we must first understand what the scientists — and engineers — are trying to accomplish with their models. For each modeling group, the question is, “How can I be certain that I have selected a correct causal relationship to effect the behavior I am interested in.” James Bernoulli (1654–1705) started search for a solution to this question [?].

Our investigation revolves around two basic issues. The first issue is the question of derivation of the experiment itself. This has to do with the approximations used and the numerical effects caused by finite numbers. The second issue deals with relative uncertainty. That is, given we know there are uncertainties in the observations, are these uncertainties enough to reverse the decision?

The basics of validation methodology were laid down by Aristotle in *Posterior Analytics*. My 1994 paper [?] discussed Aristotle’s *Organon* (part of *Posterior Analytics*) [?] and Bacon’s *Novum Organon* [?]. The two classical works are the foundation of what we now call the philosophy of science. The computer adds a new concern. I proposed the following principles for the correctness of simulations:

P1 *Physical Exactness*. We strive to identify non-physical (mathematically convenient) assumptions and eliminate them.

P2 *Computability*. We must identify non-computable relationships. Most mathematical relationships in scientific computation turn out to be approximate.

P3 *Bounded Errors*. No formulation is acceptable without a priori error estimates or a posteriori error results.

P3 is clearly a numerical methods/machine arithmetic issue and P2 is a problem for numerical analysis and computability theory. P1 represents a new way of thinking for the scientist and engineer. P1 leads us to consider confirmation as a problem addressed in the philosophy of science.
A goal of many researchers has been to develop a system of *deductive* logic that can be used to describe *inductive* science. While there are many problems with this, a related question is reasoning with uncertainty. Another position is to develop a deductive system that bridges the gap between the observational world and the formal world. I describe the Carnap-Hempel system developed over several decades. Such a system would be a good introduction to formal methods into validation.

H.1. A Quick Review of Mathematical Logic. The formalization of reasoning starts with Aristotle *Organon* (part of *Posterior Analytics*), although there was undoubtedly a long history of these investigations before he started. Aristotle’s original goal was to have a mechanism for checking arguments during debates. After the end of the “classical period,” logic languished until resurrected by, among others, Thomas Aquinas during the Scholastic era. Scholastic logic was used and improved through the 19th Century. Boole and Peirce greatly improved the notation if not the theory of logic. Peirce introduced the quantifiers, “for all” and “there exists” into common usage.

The modern version of logic can be found in any number of good texts; e.g., van Dahlen. Logic is a formal language issue (sometimes called *metamathematics*). In formal languages we describe the structure of the language and rules for manipulating the symbols of the language. In mathematical practice, there are fundamental statements that are primitive; these are the axioms. There can be many rules, but the most primitive beginning is to have *modus ponens*:

\[
\{A, A \rightarrow B\} \Rightarrow B
\]

or, in words, “From formulas ‘A’ and ‘A → B’ one can write ‘B’ alone.” From these ideas alone we can formalize *sentential logical calculus*.

There is no controversy so far. However, when we get to the quantified formulas like \( \forall x \exists y x < y \) that we get into trouble. For most situations, however, logic works the way we were taught in undergraduate mathematics.
II.2. Induction versus Deduction. Logic as we discussed above is known as *deductive* logic and this is the logic of mathematics. In deductive logic, all the information needed for the proof is given in the hypotheses. In this very important sense, no mathematical proof ever discovers anything new. In science, however, it is not the case that nothing new is discovered. In fact, in mathematics we talk about _proving theorems_ but in science we talk about _establishing laws_. A good early science course should establish in us the following distinction:

1. In mathematics, a theorem is forever. Any datum meeting the restrictions of the hypotheses will meet the restrictions of the conclusion.

2. In science, a datum that meets the restrictions of the hypotheses may _not_ meet the restrictions of the conclusion. Therefore, each experiment might fail. In fact inductive systems are contradictions of deductive systems.

These two concepts of mathematical logic and induction versus deduction were at the heart of some areas of philosophy from the mid-1930s to the mid-1970s. From the crisis in foundations in mathematics and logic from the late 1880s until Gödel’s famous Incompleteness theorem, much of philosophy and mathematics was looking for _absolute_ truth, called the _verificationist_ movement. Gödel basically showed this is an impossible goal. Even so, it is useful to have formal languages and formal rules — hence, the formal methods movement in computer science.

II.3. Axioms or Not? One of the major differences between science and mathematics is that of axioms. This actually is a question of epistemology. Mathematics answers this question in several ways, the most commonly held position is that of derivability. In the derivation scenario, the truth of a statement is shown by deriving an unbroken path from the axioms to the statement in question. This is the familiar axiomatic method.

On the other hand, science does not have this view, hewing instead to the Babylonian view. In science, computation is the key issue. The consequence of this is that the logical method developed for mathematics is not useful since there are no axioms. Feynman in [?] describes the different viewpoints and is _must_ reading.
Any formal logic that would be useful for science would require the tools to connect the axiomatic world with the computational world.

H.4. The Confirmation Problem. The confirmation problem, as the validation problem is known in the philosophy of science literature, was the focus of the philosophy of science from the mid-1930s until the mid-1970s. This focus arose from events in 19th Century science. Carnap’s 1936 and 1937 articles entitled “Testability and Meaning” are the first in a lengthy development in the logical framework of confirmation. Carnap (1891–1970), together with Carl Hempel (1905–1997) and others, developed “logical framework” for confirmation. Karl Popper (1902–1994) in The Logic of Scientific Discovery said that logical theories were set up to fail; the so-called falsification theory. Thomas Kuhn’s (1923–1996) Structure of Scientific Revolutions was among the more influential works on the conduct of science over confirmation. Kuhn tried to show that theories evolve due to confirmation/disconfirmation. In 1977 Frederick Suppe edited the proceedings to the conference The Structure of Scientific Theories: The Search for Philosphic Understanding of Scientific Theories, bringing the field into focus.

Because many computer scientists and engineers have a formal methods view, I focus on understanding the logical formalisms of Carnap and Hempel.

The goal is to formulate the conduct of science in a first-order logical system, allowing us to investigate the issues of completeness and consistency. To understand such a first-order system, we need to name the basic categories of objects and their relationships. There are three basic categories of constructs to consider: (1) the objects and the measurements, (2) science-theoretic definitions and functions, and (3) the logical axioms and operators.

For physics and engineering, we have four sorts of objects to consider: (1) idealizations, (2) inferred entities, (3) unobservable objects, and (4) theoretical postulates. These idealized objects are placed in space-time. Our functions are idealized operators over objects and space-time points; so, too, the relations. The question: how to relate observed objects to theoretical and how to understand idealized relations. That is the subject of this section.
Reduction Sentences. The basic issue is to relate these concepts: the pre-scientific meaning of a term and the ability to test or confirm the existence of the referent. Did the term æther mean anything? It would if we could confirm the existence of the referenced object. To test means having a procedure that answers a boolean question: if yes then we have a confirmatory instance and if no a disconfirmatory instance. Since absolute confirmation is impossible, we must speak in terms of degrees of confirmation. Therefore one confirmatory or disconfirmatory instance is not the whole answer.

We know that scientists and engineers do not use logic in the same way mathematicians and logicians use it. The distinction was drawn as the analytic-synthetic distinction. Analytic statements are true or false based on the form of the constituent terms. Synthetic statements are true or false based on how accurately they describe some fragment of the empirical (“real”) world. Thus, the scientific terms are synthetic and require a different approach known as a counterfactuals. Counterfactuals have several interpretations, but I prefer to think of them as primitive relations on the observations. The link between the observational world and the theoretical world is bridged by correspondence rules embodied in reduction sentences.

Our inferential structure must include statements that are synthetic as well as analytic. The concept of consequence must be defined for both synthetic or analytic consequence. In our first order language, call it $L$, derivations according to the logical rules are called $L$-consequences and those by scientific rules are called $P$-consequences. Because of this mixed mode of reasoning, we call two sentences $S_1$ and $S_2$ equipollent if $S_1$ and $S_2$ are ($L$ or $P$) consequences of one another.

The distinction is over how “implies” is translated. Computer science, logic, and mathematics all translate “implies” into material implication which is false only when the hypotheses are true and consequent is false. Counterfactuals are subjunctive implication: the false case simply is uninteresting and does not occur because it is presupposed in the antecedent. Counterfactual inference is model-based.

Then why do we need counterfactuals? Carnap [?] considers the term soluble in water: “$x$ is soluble in water” means “Whenever anything $x$ is put into water, $x$ dissolves.” Suppose we have the following:
\( Q_1(x, t) \) The body \( x \) is placed in water at time \( t \).

\( Q_2(x, t) \) The body \( x \) dissolves at time \( t \).

\( Q_3(x, t) \) The body \( x \) is soluble in water.

Our natural reaction would be to define \( Q_3 \) as

\[
Q_3(x) \equiv (\forall t)[Q_1(x, t) \supset Q_2(x, t)].
\]

But this does not capture what we want. Suppose we want \( x = b \), a glass brick. Certainly, \( Q_3(b) \) is false. But the right hand side of (18) is still true.

\[
Q_3(b) \quad (19)
\]

\[
(\forall t)[Q_1(b, t) \supset Q_2(b, t)] 
\]

Looking at (18), \( Q_1(b, t) \) must always be false and hence (19) must be true. The culprit is the form of (18). A way out of this problem is to introduce the reduction sentence. A reduction sentence has the form

\[
Q_1 \supset (Q_3 \supset Q_2)
\]

In our example, we would have the form

\[
(\forall x)(\forall t)[Q_1(x, t) \supset (Q_3(x, t) \supset Q_2(x, t)]
\]

which we could read as “If \( x \) is put into water at time \( t \), then, if \( x \) is soluble in water, \( x \) dissolves at time \( t \).”

H.6. Full Statement of the Language Requirement. Space restrictions preclude an in-depth development of the rules for developing languages that fit the Carnap/Hempel mold. See [?], pp. 50-52 for the rules and Hempel [?] for an example. I will only sketch developments here.

There are three types of statements: (1) logical, (2) observational, and (3) theoretical. The essence of the rules is that these three must be carefully woven together. There are two non-logical vocabularies: \( V_o \) for constants, functions, and relations for observational purposes and \( V_T \) for the theoretical terms of the theory.
These yield languages $L_o$ and $L_T$. The two languages are linked through correspondence rules. The correspondence rules can have two forms: (1) reduction sentences which link theoretical definitions to observation and (2) testing procedural statements. The semantics of $L$ are taken from concrete observable events, objects, or times. Therefore, the variables in $L_o$ statements must refer to variables or expressions in the observation language.

Some examples of correspondence rules that might come up in a discussion of Michelson-Morley come to mind. The definitions of place and thing-at given below are in the observational vocabulary and distance and avg-vel-between can be placed in the theoretical vocabulary. Clearly, æther is a theoretical term.

$$\text{place}(\bar{x}) \land \text{place}(\bar{y}) \rightarrow$$
$$\bar{y} - \bar{x} \text{ finite} \rightarrow$$
$$\text{distance}(\bar{x}, \bar{y}) = ||\bar{y} - \bar{x}||$$

$$\text{thing-at}(a, \bar{x}, t) \land \text{thing-at}(a, \bar{y}, t') \rightarrow$$
$$(t - t' \neq 0 \rightarrow$$
$$\text{avg-vel-between}(\bar{x}, \bar{y}) = (\bar{y} - \bar{x})/(t' - t).$$

Appendix I. Uncertainty, Sensitivity, and Numerical Issues

From a computational standpoint, we must determine the effect of using finite numerical representations for infinite ones. These measures help to quantify the uncertainty caused by calculations.

I.1. Uncertainty and Sensitivity. Some views of validation seem to focus on quantification of, and reasoning about, uncertainty as the key issue in validation. This is not just classical statistics since the statistical community has two poles: Bayesian and classical statistics. The different types can lead to different conclusions. Fortunately, we only need to quantify the uncertainty.
One measure of uncertainty is that of relative error. Let \( F(x) \) be an arbitrary, once continuous function. The \textit{relative error} of \( F \) with respect to \( x \) is defined by
\[
\text{rel. err.} = \frac{dF/dx}{F}.
\]
Since \( F \) could be multivariate, \( dF \) is the total derivative.

I.2. Sensitivity Analysis. In any theory we would expect to see mathematical relationships among the objects of the theory. When subject to observations, observational errors are introduced. A natural question would be, “How large can these observational errors be and still not effect the decision?” In other words, how sensitive is our result to errors in the observations. This is formulated by asking how does change in the error of a parameter change the function’s value. Using our arbitrary \( F(x) \) again, we want to study how \( F(x + \Delta x) \) changes with respect to the individual \( \Delta x_i \)'s.

\[
\text{sensitivity of } \Delta x_i = \frac{\partial F}{\partial \Delta x_i} \bigg|_{\Delta x_j = 0} \quad \text{where } j \neq i.
\]

I.3. Error Analysis. We must insure that the principles (see Section ??) of computational science and engineering are enforced. There are two types of errors: 1) scientific errors, denoted \( \delta_s \) and 2) numerical errors from approximations and finite arithmetic \( \delta_a \). By definition, \( \delta_s \) is unknown. Let \( F \) be the function received from the science and suppose (as a gross simplification) \( F \) is linear.

Our measure of merit for \( \delta_a \) is relative error.
\[
\text{rel. err.} = \frac{|x - \hat{x}|}{x} = \frac{dF/dx}{F(x)}.
\]
where \( x \) is the \textit{true value} and \( \hat{x} \) is the \textit{computed value}.

For the \textit{forward problem} where \( F \) and \( x \) are known and we want \( y = F(x) \), we can determine the relative error in \( y \) easily:
\[
(F + \delta_a F)(x + \delta_a x) = y + \delta_a y
\]
\[
F(x) + F(\delta_a x) + \delta_a F(x) + \delta_a F(\delta_a x) = y + \delta_a y
\]
\[
\frac{F(\delta_a x)}{F(x)} + \frac{\delta_a F(x)}{F(x)} + \frac{\delta_a F(\delta_a x)}{F(x)} = \frac{\delta_a y}{y}
\]
The first term in Eq. ?? measures the *conditioning*, the second measures the *stability* of the method and the third term measures the arithmetic errors.

1. If the conditioning number is large, then the function is extremely difficult to compute. Therefore, a different formulation of the problem is needed.

2. If the conditioning number is small, the function is easy to compute. Stability now comes into play. Stability is the measure of how approximation errors are propagated. If the stability number is large, the method probably will not be satisfactory; find another method.

3. If both conditioning and stability are good, then the errors due to finite arithmetic determine how accurate the computed value is.

Part 4. The Michelson-Morley Experiment

**Appendix J. Introduction**

The *luminiferous æther* was posited by Augustin Fresnel (1788–1827) to explain the wave properties of light. In 1886, Albert Michelson (1852–1931) and Edward Morley (1838–1923) published disconfirming results [?]. There were criticisms of the experiment. A second article appeared in November, 1887 [?]. While there was much more controversy to come, the fate of æther had been sealed. Interestingly enough, the experimental plan of Michelson and Morley was never carried to completion. Only two complete sets of observations were taken out of a year-long protocol. The æther lingered on for almost thirty years; it did not die, but simply disappeared from mainstream physics.

What does this have to do with validation of large-scale simulations? The Michelson-Morley experiment represents perhaps the finest hour of the validation methodology. The experiment had been expected to confirm, not disconfirm, the æther theory. Because the story is so well-known and understood, we can illustrate how we might apply modern ideas of validation.

**Appendix K. Illustrative Development of Æther**

This section describes the basic tenets of the Fresnel theory. This section requires only an elementary understanding of fluids and the idea of interference patterns.
To understand Fresnel’s theory, consider the æther as a fluid and light as a swimmer. If the swimmer swims back and forth between the banks (at right angles to the current), then she will be carried downstream on the two trips. When she swims with (or against) the current the speed of the river is added to (or subtracted from) to the swimmer’s speed. Therefore, there is an apparent difference in speed. This speed difference would appear as an interference pattern when this analogy is mapped over into light. The patterns can be measured quite accurately and the speed difference (the current) reflects the effect of the æther.

This section combines the developments presented in Michelson’s and Morley’s 1886 and 1887 papers.

K.1. The Formulas. Let the index of refraction $n$ be the ratio of the speed of light in the external æther ($v_e$) to that internal ($v$) to the æther. This is further defined as the inverse ratio of the square root of the densities. Suppose the density of æther is 1 outside the box but $1+\Delta$ inside a small volume (because it is moving and compressing the æther):

$$n = \frac{v_e}{v} = \frac{\sqrt{1+\Delta}}{\sqrt{1}}.$$  

We use these facts to develop the theory behind the experiment.

The question now is to determine what velocity of æther in the prism is needed to give the observed results. The speed of the æther must be

$$x = \frac{n^2 - 1}{n^2}.$$  

In terms of our observational/theoretical distinction, this entire section is done in the theoretical language. In fact, we run into trouble as soon as we try to use our observational/theoretical model.

The value of $x$ is not connected to any observable measurement: the observable $v$ is eliminated. We cannot measure the density difference between air and æther. The ratio we can measure is given by $v = v/V$. This is the key to the refutation.

We now derive the equations to support the Michelson-Morley experiment. The definitions of the variables and the assumptions are given in Figure ??.
find the time difference to predict the fringe interference pattern difference. The experimental setup is shown in Figure ??.

The times for each leg are \( T = D/(V - v) \) and \( T_1 = D/(V + v) \). The round-trip time is clearly

\[
T + T_1 = 2D \frac{V}{V^2 - v^2}. 
\]

N. B. In the 1887 paper, we have a illustration of breaking principle P1 given in Section ??: “... and the distance traveled in this time is \( 2Dv^2/(V^2 - v^2) = 2D(1 + v^2/V^2) \), neglecting terms of the fourth order (italics mine).” Having now crossed the line, the report’s authors must now continually refer to both the exact and approximate derivation.

The arm moving perpendicular to the path of the earth is actually undergoing a Lorenz contraction (we could also compute this from first principles). Therefore, the light on the other path must travel \( 2D\sqrt{1 + v^2/V^2} \). To get the approximation compatible with the above, we expand this by the binomial expansion \((a + b)^n\), with \( a = 1 \) and \( n = 1/2 \). We obtain a first-order approximation for the length as

\[
2D(1 + v^2/2V^2). 
\]

Subtracting the two distances we get \( Dv^2/V^2 \). The distance divided by the wavelength of the light is equal to an integral number of wave length; that is, the bands occur at those intervals. The experiment was run using a sodium lamp, emitting at 5,900 Å or \( 5.9 \times 10^{-7} m \).

One experiment instance is to be carried out by first setting up the apparatus one way, then rotating it 90 degrees. Therefore the total difference is twice the above, or (using the approximations), \( 2Dv^2/V^2 \). The protocol called for one reading at noon and the other at night.

K.2. The Experimental Results. As reported in the 1887 paper, observations were taken July 8–12, 1887. Measurements were taken at Noon, then the apparatus was rotated clockwise for the evening measurements. Quoting from the report:
“...It seems fair to conclude from the figure that if there is any displace-
ment due to the relative motion of the earth and luminiferous æther, this
cannot be much greater than 0.01 of the distance between fringes.

Consider the motion of the earth in its orbit only, this displacement
should be \( 2Dv^2/V^2 = 2D \times 10^{-8} \). The distance \( D \) was about 11 meters, or
\( 2 \times 10^7 \) wave-lengths of yellow [sodium] light; hence the displacement to be
expected was 0.4 of a fringe. The actual displacement was certainly less
than the twentieth part of this, and probably less than the fortieth part.

But since the displacement is proportional to the square of the velocity,
the relative velocity of the earth and the æther is probably less than one
sixth the earth’s orbital velocity, and certainly less than one-fourth.

...It appears ... if there be any relative motion between the earth and
the luminiferous æther, it must be small; quite small enough entirely to
refute Fresnel’s explanation of aberration....” [?, p280–281]

The distances were to be measured using a graduated set screw which was
reported to be able to measure 0.02 wavelengths.

**Appendix L. Applying the Principles**

We turn to the task of explaining the Michelson-Morley results in terms of the
measures we have proposed.

L.1. Can We Compute Accurately? We have already commented that the derivation
makes gratuitous use of unwarranted approximations: the time difference can be com-
puted without intermediate approximations. Without the approximations, the differ-
ence between the time along the orbit and the time perpendicular is

\[
(26) \quad \frac{2D(\rho V^2 + \rho v^2 + V)}{-V^2 + v^2}
\]

where \( \rho = \sqrt{1 + (v/V)^2} \).

How well can we compute this? One measure is known as a condition number.

The condition number is defined for a function \( f(x) \) as

\[
c = \frac{x \cdot df}{f(x) \cdot dx}
\]
If the condition number is “small” then the value of $f(x)$ is “easy” to calculate accurately. Since we have two variables, the condition number is a function of both. Using Maple, we find that (after a lot of work), that $c = 3 \times 10^{-9}$ using the values in Figure ?? . That means we should be able to compute the time difference very accurately.

Using the definition of relative error, and again using Maple, we get a relative error of $3 \times 10^{-13}$. Therefore, the computed difference should differ from the real one in the 13th decimal place. We can compute the values accurately enough to guarantee that observational errors are the dominate errors in the results.

L.2. The Logical Case On Michelson-Morley. For convenience, we copy Eq. ?? to the below and label the values as observational:

\begin{align*}
T_i &= \frac{2D_o(\rho V_o^2 + \rho v_o^2 + V_o)}{-V_o^2 + v_o^2} \\
\rho &= \sqrt{1 + (v_o/V_o)^2} \\
m_o\lambda_o &= T_i
\end{align*}

where $T_i$ is the time difference with the auxiliary definition for $\rho$. Eq. ?? is the equation relating the spacing of the interference fringes to the time difference. $m$ is actually what was calculated as 0.4 in the report on results. The report says that $D$ is “approximately 11 meters” or $2 \times 10^7$ times the wavelength of yellow light. Using the values from Fig ?? and that definition we get that $D$ is 11.8 meters. Solving Eq. ?? we find that the value of $m_t$ should be 0.3971. Taking “certainly less than a fortieth” as $1/40$, $m_o = m_t/40 = 0.00993$.

We can solve the theoretical $T_i$ equation for $v_t$. If we use the observed values and the theoretical $m$ value, we compute that

$$v_t = 29789.99999 \text{ with relative error } 0.3 \times 10^{-9}$$

(with the last two digits in error) but the observed speed in orbit would have to be

$$v_o = 4710 \text{ with relative error } 0.8 \times 10^0.$$
The zero exponent in the relative error says that the real and computed value of $v_o$
do not even agree in magnitude.

We still cannot make a connection from the observational and the theoretical
unless we accept a correspondence rule that allows

$$v_o = v_t.$$

How does this logically generate the contradiction to Fresnel’s axiom? We have both
$v$ and $V$ by measuring the speed of light in a vacuum. The value is given as $V = 2.997 \times 10^8$. If we allow such a rule then we can say that $v_o \neq v_t$ and therefore we have
a disconfirming experiment.

L.3. Sensitivity and Uncertainty. In Eq. ??, there are no arbitrary constants and just
three variables: $D$, $V$, and $v$. The sensitivity values are

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>$2.0000$</td>
</tr>
<tr>
<td>$V$</td>
<td>$3.1413 \times 10^{-15}$</td>
</tr>
<tr>
<td>$v$</td>
<td>$-2.9157 \times 10^{-11}$</td>
</tr>
</tbody>
</table>

We can conclude that we are very unsensitive since the distance is the most sensitive
but the easiest to measure: we should easily be able to measure $D$ to $10^{-3}$. The other
values are many orders of magnitude smaller, especially at the reported errors.

The relative error is $1.517 \times 10^{-16}$ as calculated in Maple with 20 digits of precision.

These numerical considerations indicate that there is low uncertainty about the
value. The sensitivity coefficients indicate that the value is quite stable within the
errors stated.

Appendix M. Lessons for Simulations

This section draws on the lessons learned above. Our task is to apply these
lessons to simulation development.

M.1. Computer Simulation, Verification, and Validation. The pragmatic development
of a computer simulation begins with the mathematical statements of the model.
While most developments treat pencil-and-paper mathematics as numerical mathematics, one must usually rewrite the mathematics. In [?], I discuss a more comprehensive view of the various forms of mathematics. The short version is (1) scientists work in terms of the mathematical analysis of infinite precision, (2) numerical analysis works in terms of convergence on rational numbers, (3) numerical methods works in terms of an idealized floating point that may or may not exist, and (4) the machine manufacturers do what they can within broad guidelines.

Several professional engineering societies have taken steps to make their members aware of these steps by developing standards. The American Institute of Aeronautics and Astronautics has worked especially hard in this area and has published “Guide to Verification and Validation of Computational Fluid Dynamics Simulations”. This is well worth reading regardless of the focus of a simulation project. The Defense Modeling and Simulation Office of the U. S. Department of Defense has extensive resources relating to best practices in verification, validation, and warranty (or by the commonly used acronym VV&A).

M.2. Specific Comments on Validation and Verification. A predominant view of validation held by many workers in the U. S. is that validation is about dealing with intrinsic uncertainty. By intrinsic uncertainty I mean the uncertainty left over after verification is completed. Based on our simple example, what is the nature of this uncertainty?

(1) Almost surely, the observations have statistical uncertainty when taken. But there could be systemic and instrument errors as well. For example, traffic counters can intermittently fail with no telltale traces of failure. This means that a simulation could be correct in the larger sense, but fail to validate since the validation data is corrupt.

(2) The validation, verification, and development groups should work together throughout the life cycle. For example, the recent loss of the Mars Polar Explorer over failure to convert MKS to British units shows how difficult it is to ferret out mistakes.
(3) Numerical errors can mask any information in the simulation. This is particularly possible for long running simulations where numerical cancellation can be a problem typified by physics codes that can run many days.

(4) The scientific explanation may be inappropriate for the problem at hand. This happens when the regime of parameters is outside the “trusted” set during extrapolation. This can occur when simulations are used outside the validated parameter space. This problems is seen as particularly difficult in large engineering simulations.

(5) The problem is so sensitive to initial conditions (like chaos) that we are unable to accurately simulate to the precision required to make a decision. Here, the simulation gives unwarranted credibility.

M.3. The Formal Methods Possibilities. The Carnap-Hempel logic provides an excellent opportunity for development of formal methods. The logical language itself is standard first order predicate calculus. The new element would be making it simple to write the counterfactual expressions and then integrating them into a whole. By the very nature of the process, it would seem quite feasible to use a logic programming language to develop a simulation writing environment.

M.4. Simulation Code Development. Development of generalized simulation writing environments requires a long lead time. This lead time might be justified for mission critical projects. But not all such projects may have the luxury to develop the logical system describing their project. This means that the simulation code developers must take up the slack.

In [?], I discussed many of the attributes needed to ensure quality simulations. Some very basic observations are:

(1) Good practice matters. Unfortunately, academics don’t teach it and organizations can’t enforce it. Software engineering practice does not address computational science and engineering.

(2) Speed must become subordinate to correctness.
(3) Detailed specifications, quality assurance procedures, and formal testing are not enough. Uncertainty by using less-well-defined algorithms is several times worse than using formal mathematical definitions.

(4) Double precision does not solve problems of unstable codes or ill-conditioned problems.

(5) Paradigm shifts in language or formal methods do not appear to automatically solve the problem. But comprehensive and objective testing, formal methods, and multiple versions may be helpful.

(6) Safe subsets for languages are very important.

(7) Construction and testing of static code fault finders are needed to
   (a) Find formally undefined behaviors in languages and systems.
   (b) Help enforce known standards.
   (c) Screen out well-defined behaviors we know we should not use.
   (d) Help assess quality.

(8) Documentation is not a panacea.

(9) Software engineering metrics and processes, as currently practiced, measure nothing of interest except those in Chapter 8 of Fenton and Pfleeger [?].

These problems still exist and are ubiquitous in systems. The point about languages goes beyond standard languages, reaching into the specialized language systems used for modeling and simulation.

M.5. The Control of Complexity in Simulations. The control of complexity of a model or simulation is an important property not seen in the Michelson-Morley experiment. The experiment was very simple, and as we have seen, not particularly prone to either observational or apparatus errors. Such is not the case with a large simulation running on several processors.

Charles Perrow, in Normal Accidents [?], advances a thesis that there are two causes for accidents such as Three Mile Island: mind-boggling component complexity and mind-boggling interconnectivity of components. We should expect there to be a point at which the complexity of the process overwhelms our cognitive ability to understand. I have called this cognitive complexity in [?]. The basic cure for cognitive complexity is traceability: the ability to trace model concepts into the the simulation.
In a large simulation, traceability is not possible without help from language tools that tie documentation to code. A beginning step would be the requirement that simulations use literate programming techniques [?]. However, documentation is not a cure-all and much more research is needed in this area.

M.6. Sensitivity. The study of sensitivity has a huge literature base in its own right. Sensitivity can be easily approached from two standpoints: control theory and operations research. In fact, these two areas are also the foundations of many simulations studies. The problem is the automatic development of sensitivity studies for simulations. The problem is exacerbated by the long run times of many simulations in use in industry. It is not uncommon to have a simulation run several days per output set.

M.7. Uncertainty and Statistics. Finally, the defining context of validation is uncertainty. Uncertainty transcends simulation type, being just as important in discrete event simulations as it is for continuous simulations. Any simulation of a stochastic system already has a great deal of statistical processing associated with the data, but this should not fool one into thinking that the uncertainty is under control. Just how robust are those statistics? How sensitive are the simulations to the actual distribution?

Among statisticians and philosophers of both mathematics and science there is a healthy debate about whether classical statistics or Bayesian statistics is correct. Practitioners have a tendency to ignore such arguments as “academic.” Unfortunately, these are not necessarily ignorable. The reason is that when all is said and done, validation is a judgment call and the judgment used is the judgment standard set by the entire intellectual community. For example, one might be roundly criticised for a judgment rendered outside best practice. Indeed, the Defense Modeling and Simulation Office of the U. S. Department of Defense has an information site at www.dmso.mil which points to the VV&A Recommended Practices Guide (RPG). A group choosing to ignore the RPG does so at its own peril.

The artificial intelligence community is the proximal source of a vast literature on reasoning with uncertainty. It is not clear exactly how this subject can be used in validation studies. It would appear to be another area ripe for formal methods.
We have presented a case study analysis of the Michelson-Morley experiment. We have presented two concepts for the validation of scientific work. One concept, the reasoning aspect, has been formalized by 20th Century philosophers of science. The reasoning aspect requires accommodation to the reasoning modes of the scientists and the careful integration of the scientific reasoning into mathematical reasoning. We saw the interplay of the observational model and the theoretical model.

The second concept concerning the ability to compute values needed in the observational and theoretical world use the ideas of conditioning, mathematical error, and convergence. We did not explore the machine’s effect through round-off error. We could do so by adding another term $\delta_r$ to Eq. ??.

The third and fourth concepts explore uncertainty and sensitivity of the final answers with respect to the errors in the problem. We found the computed values remarkably insensitive to the errors.

The task of finding workable validation techniques has hardly begun. While numerical analysis has worked out many procedures for numerical computations, most of the results are a posteriori. Understanding — indeed including — scientific reasoning in the mathematically intense theoretical model and how to logically link the observation and theoretical models is a wide open question. I present an expansion of these ideas in [?].

The question of how the lessons of the Michelson-Morley experiment can be used in simulations was partially addressed. Validation is such a direct concern that many standard ideas from areas like control theory and operations research have not been properly evaluated. I did suggest several promising approaches for possible formal methods.

Perhaps the most interesting lesson of all is that the physicists I talked look for consistency among many different values. Such scientists consider ranges of values more than a single, “correct” number. In fact, they may not be looking for a number at all: Some histories of science seem to imply that the physics world was looking for a way out of the æther for relativity; Michelson and Morley provided the exit.
Figure 4. Diagram for Eisenlohr Development

Figure 5. Schematic of Michelson-Morley Experiment

Variable Dictionary

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_e )</td>
<td>The velocity of æther outside the prism</td>
</tr>
<tr>
<td>( V )</td>
<td>Velocity of light in the medium</td>
</tr>
<tr>
<td>( v )</td>
<td>Velocity of the earth with respect to the æther; i.e., in orbit.</td>
</tr>
<tr>
<td>( D )</td>
<td>Distance between the two points of the apparatus ( ab ) or ( ac )</td>
</tr>
<tr>
<td>( T )</td>
<td>Time for light to go in direction ( a ) to ( c )</td>
</tr>
<tr>
<td>( T_1 )</td>
<td>Time for light to go in direction ( c ) to ( a )</td>
</tr>
</tbody>
</table>

Note: See Figure ?? for diagram.

Value Correspondence Rules

\[
V = (299.8524 \pm 0.0790105478190518) \times 10^6 \text{m/s} \quad \text{NIST Mickelson Project}
\]
\[
v = 2.979 \times 10^k / s = 2.979 \times 10^4 \text{m/s} \quad \text{NSSDC of NASA}
\]
\[
\lambda = 5.9 \times 10^{-7} m \quad \text{NIST}
\]
\[
v/V = 0.99 \times 10^{-4}
\]
\[
(v/V)^2 = 0.99 \times 10^{-8}
\]

Figure 6. Variable Definitions, Assumptions, and Values

To: Steve Stevenson <steve@cs.clemson.edu>
Cc: black@whitestone.ncsl.nist.gov
Subject: Re: A Case Study?
Date: Mon, 02 Nov 1998 14:25:58 -0500

Interesting paper!

I’m unsure about the argument supporting counterfactuals (3.1.1).

Just after equation 3, it says,

..., \( Q_1(b,t) \) must always be false, ...

If I understand the semantics of \( Q_1 \), that is saying the glass brick is never placed in water. Why??

Also wouldn’t the statement, "If \( x \) is put into water at time \( t \), then, if \( x \) is soluble in water, \( x \) dissolves at time \( t \)" be formalized as \( Q1 \Rightarrow (Q3 \Rightarrow Q2) \), with \( Q3 \) and \( Q2 \) switched?
## The DMSO Best Practices

<table>
<thead>
<tr>
<th>Informal Techniques</th>
<th>Static Techniques</th>
<th>Formal Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit</td>
<td>Cause-Effect Graphing</td>
<td>Induction</td>
</tr>
<tr>
<td>Desk Checking</td>
<td>Control Analysis</td>
<td>Inference</td>
</tr>
<tr>
<td>Face Validation</td>
<td>Calling Structure</td>
<td>Logical Deduction</td>
</tr>
<tr>
<td>Inspections</td>
<td>Concurrent Process</td>
<td>Calculus (unreadable)</td>
</tr>
<tr>
<td>Reviews</td>
<td>Control Flow</td>
<td>Predicate Calculus</td>
</tr>
<tr>
<td>Turing Test</td>
<td>State Transition</td>
<td>Predicate Transformations</td>
</tr>
<tr>
<td>Walkthroughs</td>
<td>Data Analysis</td>
<td>Proof of Correctness</td>
</tr>
<tr>
<td></td>
<td>Data Dependency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Flow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fault/Failure Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interface Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model Interface</td>
<td></td>
</tr>
<tr>
<td></td>
<td>User Interface</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Semantic Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structural Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symbolic Evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Syntax Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Traceability Assessment</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptance Testing</td>
</tr>
<tr>
<td>Beta Testing</td>
</tr>
<tr>
<td>Compliance Testing</td>
</tr>
<tr>
<td>Fault/Failure Insertion Testing</td>
</tr>
<tr>
<td>Graphical Comparisons</td>
</tr>
<tr>
<td>Partition Testing</td>
</tr>
<tr>
<td>Product Testing</td>
</tr>
<tr>
<td>Special Input Testing</td>
</tr>
<tr>
<td>Sub-Model / Module Testing</td>
</tr>
<tr>
<td>Visualization / Animation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compliance Testing</th>
<th>Execution Testing</th>
<th>Interface Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorization</td>
<td>Monitoring</td>
<td>Data</td>
</tr>
<tr>
<td>Performance</td>
<td>Profiling</td>
<td>Model</td>
</tr>
<tr>
<td>Security</td>
<td>Tracing</td>
<td>User Standards</td>
</tr>
<tr>
<td>Special Input Testing</td>
<td>White Box Testing</td>
<td></td>
</tr>
<tr>
<td>Boundary Value Equivalence</td>
<td>Branch</td>
<td></td>
</tr>
<tr>
<td>Partitioning</td>
<td>Condition</td>
<td></td>
</tr>
<tr>
<td>Extreme Input</td>
<td>Data Flow</td>
<td></td>
</tr>
<tr>
<td>Invalid Input</td>
<td>Loop</td>
<td></td>
</tr>
<tr>
<td>Real-Time Input</td>
<td>Path</td>
<td></td>
</tr>
<tr>
<td>Self-Driven Input</td>
<td>Statement</td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trace-Driven</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7. The DOD Best Practices List**
A little nit: the declaration of P3 on the first page has a unmatched close-quote ('' in latex). Also in Conclusions, the following line confuses me:

The second concept concerning the ability to compute values need in the observational and theoretical worlds use the ideas of ... Is that "needED" and "useS"?

Very thought provoking paper. Do you know if the metrics guys here at NIST know about this? It sounds like something they should.

-paul-

---

Paul E. Black (p.black@acm.org)  100 Bureau Drive, Stop 8970
paul.black@nist.gov  Gaithersburg, Maryland  20899-8970
voice: +1 301 975-4794  fax: +1 301 926-3696
Web: http://hissa.ncsl.nist.gov/~black/black.html  KC7PKT

------------------

From: Steve Stevenson <steve@cs.clemson.edu>
To: "Paul E. Black" <black@whitestone.ncsl.nist.gov>, steve
Subject: Re: A Case Study?
I'm unsure about the argument supporting counterfactuals (3.1.1).

Just after equation 3, it says,

... Q_1(b,t) must always be false, ...

If I understand the semantics of Q_1, that is saying the glass brick is never placed in water. Why??

Also wouldn't the statement, "If x is put into water at time t, then, if x is soluble in water, x dissolves at time t" be formalized as Q1 \implies (Q3 \implies Q2), with Q3 and Q2 switched?

I'm looking at Carnap's original paper. He writes the statement as

Q_1 \rightarrow ( Q_2 \rightarrow Q_3 )

"The description of a method of testing for ‘Q_3’ has to contain two other predicates of the following kinds:

1) A predicate, say ‘Q_1’ describing the test-condition, i.e., an experimental situation which we have to create in order to test ‘Q_3’ at a given (space-time) point.

2) A predicate, say ‘Q_2’ describing the truth-condition, with respect to ‘Q_1’ i.e., a possible experimental result of the test-condition Q_1 at a given point b of such that, if this result occurs, ‘Q_3’ is to be attributed to b.

I had a look on your entire paper again, and I like it very much. But, being neither a physicist nor a philosoph, I have some difficulties in understanding all details. It might be worth considering the typical reader of your paper and think about adding slightly