The Michelson-Morley Experiment as a Case Study in Validation

Version 2

August 6, 2002-1500

D. E. Stevenson

442 R. C. Edwards Hall

Department of Computer Science

Clemson University

PO Box 341906

Clemson, SC 29634-1906

steve@cs.clemson.edu

Abstract

We use the Michelson-Morley experiment as a case study in validation based on current knowledge. We use standard mathematical concepts of error analysis to develop measures of merit.

In order to fully understand the case study, we review philosophical concepts, focusing on scientific confirmation. We review confirmation to understand formal logical concepts that can be used. We then propose a logical foundation for scientific validation and give two research directions.

Keywords. Validation, physics, logic, numerical analysis.

I. Introduction

The Defense Modeling and Simulation Office (DMSO) is a central repository of information regarding verification and validation of models and simulations. The DMSO *Recommended Practice Guide*s [3], [4] are required reading. Therefore, a proper beginning for this work is to quote the definitions of verification and validation as given on the DMSO's web site:

- *Verification.* "The process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specification. Verification also evaluates the extent to which the model or simulation has been developed using sound and established software engineering techniques."
- *Validation.* "The process of determining the degree to which a model or simulation is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation."

Our task is to explore these concepts and propose a framework that bring meaning to *conceptual description and specification* and *real-world*.

We develop this framework by exploring a famous validation: the Michelson-Morley experiment.

> 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 [13]. There were criticisms of the experiment. A second article appeared in November, 1887 [14] that again disconfirmed the æther theory. 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 main-stream physics.

In order to understand the experiment and its implications we first give some principles for evaluating the experiment as it might be done in an interdisciplinary environment (Section II). We then describe the theoretical underpinnings and the derivation of the experiment in Section III. We apply modern analysis to the experiment in Section IV. Section IV motivates a review of epistemology in Section V, from which we develop a general approach to capturing uncertainty in Section VI. We extend the analysis to a framework for validation and propose two research directions.

II. Principles to Consider in Michelson-Morley Case Study

Understanding validation begins with understanding how we think about knowledge. We all have beliefs and the sum of all our beliefs is called a *world view*. Groups as well as individuals have world views with the group views often taking the form of professional organizations. We also hold certain ideas to exist that we call *metaphysics* or *categories*. Finally, the concept of *knowledge* comes within the categorical view. *Epistemology* is concerned with the nature, sources, and limits of knowledge. We view validation as part knowledge and part epistemology as is made clear in Section V. We provisionally define validation as "justifying" new knowledge based on the verified model and experiments in the "real world". One task in Section VI is to make *justifying* and *real-world* meaningful.

Some naïve principles from scientific and mathematical practice are possible. We distinguish between theoretical developments and observational occurrences. As discussed in [20], there is a theoretical language and an observational language which are connected through *reduction sen-*

tences and testing procedures. Reduction sentences assign meaning to theoretical symbols as in Figure 4 while testing procedures relate to such things as acceptable statistical procedures on data.

In [19], I listed three categories of players in computational science and engineering: application researchers (science, etc), algorithm researchers, and architecture researchers (software and hardware). Each brings her/his own world view, categories, knowledge, and epistemology. For the Michelson-Morley, we need all three: physics for the meaning, mathematics to derive the expressions, and computation to evaluate the expressions.

Mathematics uses the term *error* to indicate the difference between the true and calculated solutions. [19] also lists three principles required for models:

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.

We note that the errors generated by computation must be less than the uncertainty in the application's values. These principles are further explored in Section V.

A. Conditioning and Error Analysis

The term *condition* describes a measure our ability to compute a value of a problem $y = F(x)$, any or all of which might be vectors. If the condition number is "small" then the value of $F(x)$ is "easy" to calculate accurately. Condition numbers were introduced by Wilkinson [24] with the generalities relying on tensors. Briefly, let $x_1, x_2, \ldots, x_m$ be parameters to a problem and

$y_1, y_2, \ldots, y_n$ be the solution. The mn terms

$$k_{rs} = \frac{\partial y_r}{\partial x_s} \; (r = 1, \ldots, m, s = 1, \ldots, n)$$

is the complete information on the sensitivity of the solution with respect to perturbations of the parameters. Because most problems have mn large, various approximations to the condition of the problem are used. The Michelson-Morley experiment has few variables so conditioning analysis can be carried out in full.

We are now in a position to study how one might use the principles **P1–P3** above. We consider two types of errors: 1) scientific errors, denoted δ_s and 2) numerical errors from approximations and finite arithmetic δ_a. From the mathematical perspective, δ_s is unknown and consequently ignored. Let F be the function received from the science and suppose (as a gross simplification) F is linear.

One measure of merit for δ_a is *relative error*:

$$\text{rel. err.} = \frac{|y - \hat{y}|}{y} = \frac{dF/dx}{F(x)}.$$

where x is the *true value* and $\hat{x}$ is the *computed value*.

For the *forward problem* where F and x are known and we want $y = F(x)$, we can determine the relative error in y:

$$(F + \delta_a F)(x + \delta_a x) = y + \delta_a y \tag{1}$$

$$F(x) + F(\delta_a x) + (\delta_a F)(x) + (\delta_a F)(\delta_a x) = y + \delta_a y$$

$$\frac{F(\delta_a x)}{F(x)} + \frac{(\delta_a F)(x)}{F(x)} + \frac{(\delta_a F)(\delta_a x)}{F(x)} = \frac{\delta_a y}{y} \tag{2}$$

The first term in Eq. 2 measures the conditioning, the second measures the *stability* of the method and the third term measures the *finite arithmetic errors*. We interpret Eq. 2 as:

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.

B. *Uncertainty and Sensitivity*

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. Clearly, one measure of uncertainty is that of relative error.

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 Δx_i's.

$$\text{sensitivity of } \Delta x_i = \left. \frac{\partial F}{\partial \Delta x_i} \right|_{\Delta x_j = 0} \qquad \text{where } j \neq i.$$

Some final comments. First, we have presented these measures in their simplest fashion, what one might call first order analysis. Clearly one can expand these ideas; e. g., conditioning can be carried out in higher order derivatives. Secondly, these concepts are known by different names is different disciplines; e. g., in economics, conditioning is known as *elasticity*.

III. THE MICHELSON-MORLEY EXPERIMENT

This section describes the Michelson-Morley experiment, starting with the basic tenets of the Fresnel theory. This section requires only an elementary understanding of fluids and the idea of interference patterns. The development here follows Eisenlohr as described in [23]. This section combines the developments presented in Michelson's and Morley's 1886 and 1887 papers.

As a metaphor for understanding the model, consider the æther as a fluid and light as a wave as shown in Figure 2. If the wave travels back and forth between two reflectors (at right angles to the current), then it will cover the distance at some apparent, constant speed. When it travels with (or against) the current the speed of the river is added to (or subtracted from) to the wave's speed. Therefore, there is an apparent difference is speed. This speed difference would appear as an interference pattern. The patterns can be measured quite accurately and the speed difference (the current) reflects the effect of the æther.

A. *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 (Figure 3). 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):

$$n = \frac{v_e}{v} = \frac{\sqrt{1+\Delta}}{\sqrt{1}}. \tag{3}$$

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

$$x = \frac{n^2 - 1}{n^2}. \tag{4}$$

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. 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 4. We need to find the time difference to predict the fringe interference pattern difference. The experimental setup is shown in Figure 2.

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 II. "... 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 5,900Å 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.

B. 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 displacement 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 2×10^7 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...." [23, p280–281]

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

IV. Applying the Principles

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

A. Can We Compute Accurately?

We have already commented that the derivation makes gratuitous use of unwarranted approximations: the time difference can be computed without intermediate approximations. This was done with the Maple derivation. Without the approximations, the difference between the time along the orbit and the time perpendicular is

$$\frac{2D(\rho V^2 + \rho v^2 + V)}{-V^2 + v^2} \tag{5}$$

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

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

$$C_f^x = \frac{x}{f(x)} \frac{df}{dx}$$

Using `Maple`, we find that (after a lot of work), that the the condition number is 3×10^{-9} when using the values in Figure 4. 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 3×10^{-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.

B. The Logical Case On Michelson-Morley

For convenience, we copy Eq. 5 to the below and label the values as observational:

$$T_i = \frac{2D_o(\rho V_o^2 + \rho v_o^2 + V_o)}{-V_o^2 + v_o^2} \quad (6)$$

$$\rho = \sqrt{1 + (v_o/V_o)^2} \quad (7)$$

$$m_o\lambda_o = T_i \quad (8)$$

where T_i is the time difference with the auxiliary definition for ρ. Eq. 8 is the equation relating the spacing of the interference fringes to the time difference. m_0 is actually what was calculated as 0.4 in the report on results. The report says that D is "approximately 11 meters" or 2×10^7 times the wavelength of yellow light. Using the values from Fig 4 and that definition we get that D is 11.8 meters. Solving Eq. 8 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

$$v_t = 29789.999\underline{99} \text{ with relative error } 0.3 \times 10^{-9}$$

(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.

C. Sensitivity and Uncertainty

In Eq. 5, there are no arbitrary constants and just three variables: D, V, and v. The sensitivity values are

Variable	Value
D	2.0000
V	3.1413×10^{-15}
v	-2.9157×10^{-11}

We can conclude that we are very *un*sensitive 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 1.517×10^{-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.

V. GENERALIZATION

If computational science and engineering were only about physics then we would be able to draw specific rules about validation from this experiment. But computational science is much broader and as much about modeling as it is about computation. We must develop general guidelines that makes modeling and validation part of the same process.

A. Unified Conduct

We consider the conduct of the whole experiment paying particular attention to the use of knowledge.

1) Using current knowledge, we derive a model.
2) Using the proposed model, we can derive implications of that model that would be consistent with current knowledge.
3) Using the implications, we perform experiments that result in evidence for and against the new model.
4) Comparing the theoretical implications and the observations we arrive at a conclusion concerning the coherence of the model with current understanding.
5) Taking the changes to current knowledge arrived at above, we attempt to integrate the new knowledge into the old.

Items 1–2 are formal, logical steps using deductive arguments based on the current state of knowledge. That knowledge base is assumed to be consistent, at a minimum. Item 5 is the crux of the validation process. The new knowledge must be fit into the old knowledge base, a process that is often difficult scientifically and in human terms since it must be accepted by the community. We conclude that the usual slogan about validation is not correct: validation is not “Did you build the right thing” but rather “Can you fit the model into accepted knowledge bases using the allowed epistemological processes?”

The point is that validation is about knowledge and fitting ones experiences into current knowledge. An interesting take on this process is [5]. We turn now to a micro-description of knowledge and epistemology.

B. Philosophy in a Nutshell

Knowledge is defined as "justified, true belief" or "warranted belief". Knowledge is consensual and communal and not absolute. The community sets the limits of uncertainty. We generally mean *propositional knowledge* or *explicit knowledge* [17]. *Warrant* is both a noun and a verb (see the *Oxford English Dictionary*). The verb *warrant* is a synonym for *certify* but its technical meaning is to add something to experience to form knowledge.

Epistemology is rules for turning experience into knowledge. Warranting is only part of epistemology. This section suggests how the philosophical concepts might be implemented.

1) Knowledge: Knowledge is belief held to be true by a community. There are three major issues:

1) What ought we believe or *implicit knowledge*,
2) What types of reasoning should be allowed in order for knowledge to be warranted, and
3) What necessary and sufficient conditions for knowing which and why reasons in 2 hold.

2) Epistemology: Several overarching rules have been proposed, of which *correspondence* and *coherence* are two standard rules. Several other specific rules are given in Section VI.

Correspondence rules fit the following schema. A sentence is *true* if and only if the sentence corresponds to a state of affairs and that state actually occurs. This leads to considerations of information and evidence. In the Michelson-Morley case, the experimental values evaluated in the theory do lead to correct results.

Coherence has degrees. Coherence relates to the degrees of interconnectedness or unification of the knowledge. There are three standard concepts:

1) Logical coherence: consistency.
2) Inferential coherence: soundness and probabilistic measures.

3) Explanation: relevance.

We can use these concepts to judge the value of the confirmation concepts proposed. The experiment certainly fails the last point, since it does not explain the process.

Without further discussion, we accept here the current confirmation theory approach to science.

C. Confirmation Theory

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. Carnap's 1936 and 1937 articles entitled "Testability and Meaning" [2] 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* [16] said that logical theories were set up to fail; the so-called *falsification* theory. Thomas Kuhn's (1923–1996) *Structure of Scientific Revolutions* [12] was among the more influential works on the *conduct* of science over confirmation. Kuhn tried to show that theories evolve due to confirmation/disconfirmation.

The central focus of confirmation is the belief that science is lawful. 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 operators are idealized operators and relations over objects and space-time points. The question: how to relate observed objects to theoretical and how to understand idealized relations.

Any statement made about a system could be true, false, or neutral. But this valuation could vary over time: for example, better instruments and better measurements might resurrect the æther. There are two problems: (1) how to evaluate hypotheses over time in light of an increasing amount of evidence and (2) when is enough evidence enough. In other words, the warranting process appears to be a type of decision problem. One way this could be done is discussed in [22].

D. The Reasoning and Risks

More specific rules other than correspondence and coherence have been proposed. For example, DMSO [3], [4] proposed that validation concern itself with credibility, capability, and relevance with a focus on fidelity (accuracy, precision, and sensitivity). Typical scientific studies have inherited rules such as the four sigma rule in statistical inference.

The DMSO credibility criterion needs further amplification. In particular, many verification and validation efforts do not pay enough attention to the role people play. For this reason, we emphasize that knowledge is held by people and that people bless rules and people can make mistakes. Recall two such incidents: The *Challenger* disaster in 1986 and the Cerro Grande fire at Los Alamos in 2000. In each case, people misread the evidence.

Therefore, any validation approach must face the essential uncertainty of the enterprise and the frailty of people.

N. B. A clear necessity is measurements but that is necessary but not sufficient. We must reason about the system based on these measurements and we should not think that they are a cure-all.

E. Confirmation and Explanation

A useful standard is that a validation should insure that the model explains the phenomena it models. Explanation is a logical concept that has the following idealistic component regarding prediction:

1) If a law is approximately true, then we should expect its predictions to be approximately true.
2) If a law is approximately not true, then we should expect its predictions to be approximately not true.

In symbols we might say something like

$$(L \text{ and } P) \text{ or } (\neg L \text{ and } \neg\, P)$$

The catch is the word *approximately*. The problem is solved by *counterfactuals* [20] with the following schema: C explains E if

1) C and E be events.
2) C be prior to E.
3) If C had not occurred E would not have occurred, all else be equal and
4) If C had occurred in similar situations, E would have occurred.

This leads us to a mode of reasoning known as *denying the consequent*:

$$\frac{L \to P \quad \neg P}{\neg L}$$

which is backwards to the usual forward direction of *affirming the antecedent*[1].

$$\frac{L \to P \quad L}{P}$$

[1] In the logic and mathematical literature, affirming the antecedent is known as *modus ponens* and denying the consequent is *modus tollens*

VI. A Framework for Integrated Verification and Validation

A. Formal Methods

Formal methods are defined as "mathematical approaches to software and system development which support the rigorous specification, design and verification of computer systems" by Formal Methods Europe (www.fmeurope.org). Formal methods are used by NASA, for example, in the development of a collision avoidance system *Airborne Information for Lateral Spacing (AILS)*. In the pure computer science context, again quoting FM Europe,

> The use of notations and languages with a defined mathematical meaning enable specifications, that is statements of what the proposed system should do, to be expressed with precision and no ambiguity. The properties of the specifications can be deduced with greater confidence and replayed to the customers, often uncovering facets implicit in the stated requirements which they had not realized. In this way a more complete requirements validation can take place earlier in the life-cycle, with subsequent cost savings.

Our goal is to develop formal methods for scientific simulations.

B. The Paradigm

In Section V we considered validation as a decision process: the decision is whether or not the model fits into the knowledge base of the application as augmented by mathematical and computational considerations. We now present a possible formulation.

The first issue is to decide what we will try to fit into the knowledge base. In light of our focus on knowledge, we begin by realizing models are formulated to answer questions. Models are then developed to study various facets of the question. In the Michelson-Morley experiment, the question was the ontology of the Fresnel theory. Using all the information (knowledge) at hand, a model was derived that allowed theoretical conclusions about observations. In our case, the measures of merit were the the observables. The issues of which experimental values are mapped to which theoretical variable and how the values are manipulated was settled by the correspondence rules.

The experiment was run and each replicate becomes a warrant. The warranting process is to determine how the values observed match the known (and accepted!) values. In our case, the results were disconfirming rather than confirming.

C. *Verified Model Formalization*

Let $\mathbb{Q}$ be the logical statement of the questions to be answered. The current knowledge base $\mathbb{K}$ is a system of logical statements. The initial question is "Is the question consistent with the knowledge base?" In symbols: $\mathbb{K} \rightarrow \mathbb{Q}$.

Simulations as computations can be logically expressed in *Hoare logics* [1]. If the input specification of the simulation is $\mathbb{S}_{in}$, the processing as S, and the output specification is $\mathbb{S}_{out}$, then we denote the logical statement that "Given the input state specification $\mathbb{S}_{in}$, the computation S terminates in the state $\mathbb{S}_{out}$" as

$$\mathbb{S}_{in}\{S\}\mathbb{S}_{out}.$$

Therefore,

$$\mathbb{K} \rightarrow \mathbb{Q} \rightarrow \mathbb{S}_{in}\{S\}\mathbb{S}_{out}$$

What we normally call a model is the derivation of the application ($\mathbb{D}$) to the questions. Therefore, denote the *unverified* model as

$$\mathbb{M} = (\mathbb{K} \rightarrow \mathbb{Q} \rightarrow \mathbb{S}_{in}\{S\}\mathbb{S}_{out}, \mathbb{D}_{\mathbb{M}}). \tag{9}$$

Let $v\mathbb{M}$ denote the verified verified model. Logically, this is the model, the derivation, and a proof that the derivation is correct:

$$v\mathbb{M} = (\mathbb{M}, \text{Proof}(\mathbb{D})).$$

D. The Decision Formalization

The questions that led to the verified model are developed into a decision process by another derivation. The decision model is written similarly to Eq 9:

$$v\mathbb{DP} = (v\mathbb{M}, \mathbb{D}_{\mathbb{DP}}). \tag{10}$$

The measures of merit mm are defined by $v\mathbb{DP}$. This verified decision process is formed from the logical conclusions of the verified model. The decision process *is* the warranting process for this particular verified model.

Given the verified model and the correspondence rules, we can define the observations and the theoretical versions of the measures of merit. Let cd be the values incorporated into the verified model by the reduction sentences and md be the values from the solved model. A warrant is one triple (mm, cd, md). Therefore the decision is made of the set of warrants available at the time the decision is made.

The warranting process is defined by explanations that are taken to posses the properties shown in in Figure 1. That is, the warranting process must insure that the new knowledge fits the old knowledge in at least those properties.

E. The Probability Knowledge Graph

All knowledge bases can be depicted as a forest of trees, not necessarily rooted nor just one tree. A proof is a tree. Suppose we have a proof tree (actually, any verification tree should work) and that we are able to assign probabilities to some nodes. Then a linear program can be used to

Coherent	Law-Like
Consistent	Organized
Credible	Predictive
Justified	Reasoned
Relevant	

Fig. 1. Properties of Explanations

assign a consistent set of probability intervals to the rest of the trees[8]. Call this processed tree a *probabilistic knowledge graph* or PKG. An important property of the PKG is its existence.

1) If the tree cannot be built, then the warrant is disconfirming.
2) If the tree can be built,
   a) But lowers critical measures, then the warrant is ambiguous.
   b) And raises critical measures, then the warrant is confirming.

This process can be repeated as long as resources are available.

In terms of the warranting properties, the following facts exist:

1) Coherence, Consistency, Justification, Lawfulness, Organization, Reasoned content are all inherent to the justification process.
2) Prediction comes from the measures of merit.
3) Relevance comes from the appearance of particular item in the PKG

We can now comment on the DMSO criteria. DMSO lists credibility, capability, relevance, and fidelity as its properties. These properties are covered by the above except for credibility.

Credibility is not a mathematical property but a human decision based on intangible factors. One model of credibility is evaluate the following factors: risks, people, budget, time, and quality. For example, if the project has high risk, poorly-trained people, low budget, and short time frame there is a high probability that the quality of the effort is low and hence no credibility. Our only rule for this model is that the credibility factor is an optimization, not a deterministic process — managers cannot set more than three of the values.

F. Using the PKG

The PKG, in and of itself, does not answer validation requirements because it captures only one situation. The PKG captures the relationship between the theoretical language and one set of observations. We expect many experiments to be necessary. How do we arrive at a decision given these multiple experiments?

Two approaches come to mind. The first is a formulation of Dempster-Shafer theory based on uncertain dynamical systems. The second approach follows Polya and Jaynes that formalizes evidence as log odds. Each approach is wide open to further research.

1) Shafer Structures: The Dempster-Shafer formalism is widely used to deal with systems with uncertainty [18] using axiomatic definitions of belief functions. Briefly, Dempster-Shafer theory posits *basic probability assignments*, *belief* or *support* functions, and *plausibility* functions. The latter two are defined on basic probability assignments. Support functions measure the support of a hypothesis for a conclusion, while plausibility functions measure suitability. Kohlas [11] proposes a Dempster-Shafer system relating to *uncertain dynamical systems*. This formulation is attractive since many models are of dynamical systems. We describe enough of Kohlas' work to give a flavor of the approach.

Let Θ be a set of possible values (the resolution set) for variable X. Suppose that there is conflicting evidence as to which interpretation of X is correct. Let Ω be the set of possible interpretations. Let $\omega \in \Omega$ be an interpretation and $\Gamma(\omega) \subseteq \Theta$ is the set of values X can would take on were ω be the true interpretation. In a probability setting, this situation induces a probability space

$(\Omega, \mathcal{A}, P)$ where the probability measure P describes the likelihood of the different interpretation. Define *hints* as

$$\mathcal{H} = (\Omega, \mathcal{A}, P, \Gamma, \Theta).$$

Hints are interpreted as uncertain restrictions on X: $X \subseteq \Gamma(\omega)$.

Let $H_X \subseteq \Theta$. The question is, "To what degree is H_X supported by $\mathcal{H}$?" At this juncture, Kohlas generates two concepts in Shafer: support and plausibility.

1) The *degree of support* $sp(H)$ is the probability that a particular interpretation ω is the correct one:

$$sp(H) = P(\{\omega \in \Omega : \Gamma(\omega) \subseteq H\}).$$

2) The *degree of plausibility* measure the probability that H is not excluded

$$pl(H) = P(\{\omega \in \Omega : \Gamma(\omega) \cap H \neq \emptyset\}).$$

This formulation can be used to complete the development of Shafer-like systems. But it has uses in dynamical systems. Let $\mathcal{X} = \{X_i\}$, a sequence of variables and $\Theta_i = \Theta_0$ for all i, be the state space. For uncertain dynamical systems, we can describe the uncertain transitions as

$$\mathcal{T}_i = (\Omega_i, \mathcal{A}_i, P_i, \Theta_{i-1} \times \Theta_i)$$

Uncertain transitions represent uncertain inclusions; *i. e.*, what elements are in the next state. We can similarly define uncertain observation relations. With this, control-theoretic formulations of systems are possible.

2) Information Approach: We now turn to a model of evidence accumulation, building on the works of [6], [7], [9], [10], [15]. Information is a measure of our knowledge of the state of the system after an event relative to our overall knowledge of the state. Information is a measure of relevance. (This argument first appeared in [21].)

We first consider the *likelihood* function:

$$L(H : C|G) = \frac{P(C|H\&G)}{P(C|G)}, \tag{11}$$

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 *information* leads to the interpretation that if the likelihood is one, we would say there is no information:

$$I(H : C|G) = log_b L(H : C|G). \tag{12}$$

The base b of the logarithms is immaterial allowing us to use "natural" units, such as decibels and bits.

Following [6], [7], 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,

$$\begin{aligned} W(H : C|G) &= I(H : C|G) - I(\bar{H} : C|G) \\ &= \log_b \frac{P(C|H\&G)}{P(C|\bar{H}\&G)} \qquad (13) \\ &= \log_b F(H : C|G), \qquad (14) \end{aligned}$$

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*

$$
\begin{aligned}
e(H : C|G) &= e(H|G) + \log F(H : C|G) \\
&= e(H|G) + W(H : C|G).
\end{aligned}
$$

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.

A clear problem with this formulation is that we do not have a single variable but a vector of probabilities. If the matrix used to generate the probabilities were square, then we could solve the defining equation for logarithm. Since matrices are unlikely to be square, a *pseudo-logarithm* is needed.

VII. Summary and Conclusions

We have explored the Michelson-Morley experiment using mathematical and statistical concepts not used at the time of the experiment. In order to generalize the lessons of Michelson-Morley, we reviewed philosophical concepts of world view and epistemology. In interdisciplinary endeavors we can anticipate a need to merge epistemological views of the team.

Based on these philosophical ideas we then developed a generalized view that was used to formulate a formal model of the validation process. The verification of the model serves as a mechanism for assigning probabilities to inferential statements: the probabilistic knowledge graph (PKG). We proposed two uses for the PKG: one using a variant of Dempster-Shafer theory and the other using the evidential view of Jaynes.

We conclude with answering the tasks set out in the introduction: what do the terms *conceptual description and specification* and *real-world* mean. We take

- The *conceptual description and specification* to be our verified model and

- The definition of *real-world* to be the PKG.

But saying this does not make it so and hence we have raised more questions than we have provided answers for.

ACKNOWLEDGMENTS

Published papers are often as much the work of the referees as the authors. I would especially like to thank one referee who invested a great deal of time in commenting on every aspect of the original draft. This new version is much different in its orientation and message.

REFERENCES

[1] Krzysztof R. Apt and Ernst-Rüdiger Olderog. *Verification of Sequential and Concurrent Programs*. Spring-Verlag, 1991. QA76.76.V47 A67 1991.

[2] R. Carnap. Testability and meaning. *Philosophy of Science*, 3 and 4, 1936–1937.

[3] DMSO. Vv&a recommended best practices guide. US Department of Defense, November 1996.

[4] DMSO. Vv&a recommended best practices guide. US Department of Defense, May 2000.

[5] Michael Gibbons, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. *The new production of knowledge : the dynamics of science and research in contemporary societies*. SAGE Publications, Thousand Oaks, CA, 1994.

[6] I. J. Good. *Probability and the Weighing of Evidence*. C. Griffin, London, 1950. QA273.G64.

[7] I. J. Good. *Good thinking : the foundations of probability and its application*. University of Minnesota Press, Minneapolis, 1983. QA273.4 .G66 1983.

[8] Theodore Hailperin. Probability logic in the twentieth century. *Hist. Philos. Logic*, 12(1):71–110, 1991. MR 92j:01044.

[9] E. T. Jaynes. *Maximum-Entropy and Bayesian Methods*, chapter Probability Theory as Logic. Kluwer, 1990.

[10] Harold Jeffreys. *Theory of Probability*. At The Clarendon Press, Oxford, 3rd edition, 1939.

[11] Jurg Kohlas and Paul Andre Monney. *A mathematical theory of hin Shafer theory of evidence*. Springer, 1995.

[12] Thomas S. Kuhn. *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago, IL, 2 edition, 1970.

[13] Albert Michelson and Edward Morley. Influence of motion of the medium on the velocity of light. *American Journal of Science*, 31:380, May 1886. 3d series.

[14] Albert Michelson and Edward Morley. On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, 34:333–336, Nov 1887.

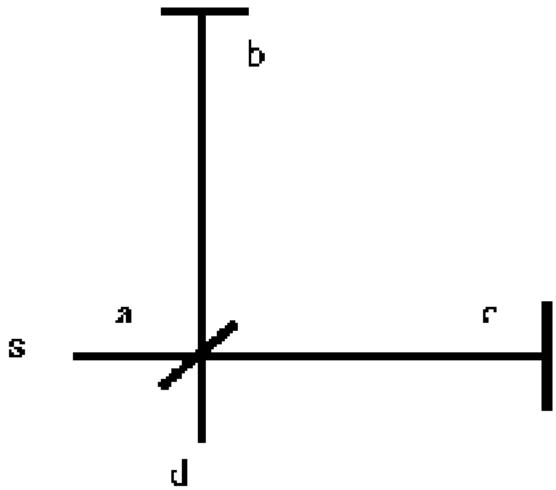

Fig. 2. Schematic of Michelson-Morley Experiment

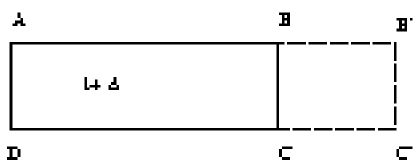

Fig. 3. Diagram for Eisenlohr Development

[15] George Polya. *Mathematics and Plausible Reasoning*. Princeton University Press, 1968.

[16] Karl Popper. *Logik der Forschung/The Logic of Scientific Discovery*. Basic Books, New York, 1935 (1959).

[17] Gilbert Ryle. *The Concept of Mind*. Hutchinson, 1949.

[18] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[19] D. E. Stevenson. Science, computational science, and computer science: At a crossroads. *Comm. ACM*, 37(12):85–96, 1994.

[20] D. E. Stevenson. A foundations of validation: The michelson-morley experiment. In *The Proceedings of the European Simulation Multiconference 1999, Warsaw, Poland, June 1-4, 1999*, pages 269–275, 1999.

[21] D. E. Stevenson. An evidence-based approach to fidelity. In *Proc. SCSC 2001, San Diego, CA, 16–18 Jul 2002*. SCS, 2001.

[22] D. E. Stevenson. An evidence-based approach to fidelity. In *Proc. SCSC '01, Orlando, FL, July, 2001*, 2002.

[23] Loyd S. Swenson, Jr. *The Ethereal Aether*. Univ. of Texas Press, Austin, TX, 1972.

[24] J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Wiley, New York, 1963.

Variable Dictionary

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

Note: See Figure 2 for diagram.

Value Correspondence Rules

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

Fig. 4. Variable Definitions, Assumptions, and Values

Assertion Checking	Comparison Checking	Execution Testing
Fault/Failure Testing	Field Testing	Functional Testing
Predictive Validation	Regression Testing	Sensitivity Analysis
Special Boundary Testing	Statistical Testing	Structural Testing
Sub-Model Testing	Top-Down Testing	

Fig. 5. Validation Usable Techniques