KEY TERMS AND CONCEPTS

The two basic terms for this paper are *fidelity* and *validity*, along with their associated referents (the standards by which those are measured). In order to deal correctly with fidelity and validity, we also have to address the idea of *uncertainty*, which has both *error* and *stochastic variation* components. To put these terms and concepts in perspective, we present a paradigm that shows how system theory (our understanding of reality) is developed, how a simulation is developed, the relationship between system theory and simulation developments, and how V&V relates to them.

Although fidelity and validity are fundamental concepts in model and simulation (M&S) verification, validation, and accreditation (VV&A), neither their definitions nor their referents have become as precise or enjoyed as widespread acceptance as the definitions of verification, validation, or accreditation. This is one reason that so little progress has been made in our ability to quantify M&S fidelity and validity. We use the definitions below.

Fidelity

The degree to which a model or simulation reproduces the state and behavior of a real world object or the perception of a real world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness This is essentially the definition of fidelity developed by the fidelity study group in the Simulation Interoperability Workshop (SIW), [1] & [2]. Fidelity should generally be described with respect to the measures, standards or perceptions used in assessing or stating it. It will be noted that fidelity is an absolute measure of M&S representational closeness to reality. The fidelity of representation does not change with application demands. The M&S glossary on the Defense Modeling and Simulation Office (DMSO) web site defines fidelity as “the accuracy of the representation when compared to the real world.” Accuracy is usually associated with a parameter, and usually does not imply both the scope of representation (the factors represented) and the quality of representation of those factors (which is where “accuracy” is normally applied); therefore, it seems better to use the definition above.

Validity

The quality of being inferred, deduced or calculated correctly enough to suit a specific application, with particular application to a model or simulation’s representational capability. The logical truth of a derivation or statement, based on a given set of propositions. It will be noted that validity is a relative measure of M&S representational closeness to reality; validity determines whether the representation is close enough for an intended purpose. The validity of representation changes with application demands. Thus, while the fidelity of an M&S’ representation is constant, it could be valid for one application and not valid for another. This extremely important concept is often misunderstood. The definition for validity in the M&S glossary on the DMSO web site is concerned only with data quality and is not appropriate for this paper.

Referent

A codified body of knowledge about a thing being simulated. It is a standard from which the thing being represented in simulation is derived or the standard against which the correctness of M&S representation is measured (this is essentially the definition of referent developed by the SIW fidelity study group, [1] & [2]; the M&S glossary on the DMSO web site does not include a definition for referent). The *fidelity referent* is the most complete collection of information about the subject represented in the M&S; it is basically the system theory in the paradigm below. The *validity referent* is more complex because it must involve intended use for the M&S. This can cause parts of system theory and test/experimental data to be ignored in the validity referent. For example, an M&S for a beginning physics course might ignore friction in order to illustrate principles of momentum and kinetic energy more clearly to students. Such approaches, though valuable educationally, can complicate selection of what is appropriate for the validity referent. Some suggest that the conceptual model for the simulation (if properly done) is the best generic validity referent since it captures both pertinent elements of system theory and intended M&S application.

Unfortunately, no widely accepted set of terms yet exists for definition of uncertainty, error, and variability as they apply to simulation fidelity, validity, and their referents (which includes the science and test foundation upon which they rest). There are numerous variations in the ideas about these. More than half the papers expected for the V&V Symposium at the US National Congress on Computational Mechanics (August 2001) address simulation uncertainty and errors. Some in the M&S community follow the approach of the National Institute of Standards and Technology (NIST) in saying that uncertainty has two components, one that can be described statistically and one that must be described by other means, [3]. Others (such as some working on ASCI V&V at Sandia National Laboratories and the AIAA *Guide for V&V of CFD*, [4]-[7]) separate error (which they define as uncertainly caused by known or knowable factors) from uncertainty (which they define as caused by inherent variability of the process or item under consideration). Yet others (Jaynes, [8]) would deny the existence of stochastic variation when things are considered at their logical limit, which introduces a consideration that most people ignore: uncertainty that derives from the limits of logic. We have chosen simple dictionary-like definitions for uncertainty, error, and stochastic variation.

Uncertainty

Not known reliably. This means that one cannot act with assurance on information or data in a region of uncertainty. Uncertainty, if it can be quantified, establishes a boundary on the accuracy with which something can be known. All observations and measurements include some degree of uncertainty that result from a combination of error and variability. Such uncertainty may be imposed by resolution limits in the observing device, bias, lack of complete knowledge about the environment of the observation, etc. At some level, the very act of measurement introduces uncertainty, as expressed in the Heisenberg Uncertainty Principle. National Institute for Standards and Technology (NIST) Technical Note 1297, [3], emphasizes the importance of quantifying uncertainty related to measurement in order for information from that measurement to be useful, an important and much needed emphasis because some data used to develop system theory or data which may be used as M&S fidelity and validity referents fail to include quantification of uncertainty relative to the data.

Less well known by many in the M&S community, is uncertainty imposed by logical limitations. Mathematicians and logic specialists know Kurt Goedel’s 1931 treatise on formally undecidable propositions, which demonstrated that statements could be made about the real number system that cannot be proven true or false (even when one has complete and perfect knowledge). This imposes a logical limit on certainty of results from simulations. Simpson’s Paradox is a term for spurious correlations when lurking variables which impact conclusions have not been identified, such as may occur when conclusions based upon indications by the elements of a set are at odds with conclusions from the set as a whole. This potential problem has significant implications for "piecemeal validation". As a final example, consider the Cantor Point Set in infinite set theory which allows operations (such as throwing a line segment away completely while retaining it completely) that are impossible in finite mathematics; some of the approximations used in computationally intensive simulations may depend upon premises in infinite mathematics to generate finite solutions without due consideration to implications for consequent uncertainty in results.

Error

Difference between the true value and an observed or computed value. Error is the inverse of accuracy. This definition does not attempt to decompose error into its various components, nor to separate what can be known (often called "systematic error" or "bias," and presumably reducible by appropriate efforts such as repeated tests and improved calibration) from the inherent variability in what is being measured. Errors caused by variability of what is being observed are sometimes called "sampling errors". Errors associated with referents as well as errors in simulation results must be quantified explicitly for simulation fidelity and validity to be quantified usefully.

Stochastic Variation or **Variability**

The inherent variation in the process, item, system, or environment under consideration. Some also use the term stochasticity for this variability. Often the variation is random (or presumed to be random) in that the current instance of the parameter under consideration is independent of relationship to previous instances of that parameter (such as the likelihood of a head or a tail in flipping a coin does not depend upon previous flips of the coin, if done honestly). Typically when dealing with topics above the quantum level, such variation requires that the process, item, system, or environment be described in statistical terms in order to represent its variability. M&S representation of such parameters usually involves distribution functions, lookup tables, etc. With enough information about variability, useful description about uncertainty in the process, item, system, or environment resulting from that variability can be developed.

PARADIGM FOR SYSTEM THEORY, SIMULATION DEVELOPMENT, AND V&V

Figure 1 shows how system theory develops, drawing upon what is known from experiment, test, observation, and data, and then expanding that understanding though hypothesis to identify new and additional tests and data needed to confirm (or negate) elements of the theory. Note that we use "system theory" as *the shorthand for all organized knowledge and information about a topic*. Some hypotheses used to describe system data and to extend theory beyond the bounds of extant data are based upon simulation results. This figure also shows how simulation development progresses and where V&V fits. It shows how system theory, simulation development, and V&V interact. The combination of system theory and system data serves as the fidelity referent. Since validation is relative to intended use, the validity referent must temper the fidelity referent by intended use, which makes the conceptual model the primary validity referent since the simulation conceptual model is a simulation developer's way of translating modeling requirements into a detailed design framework, from which the software, hardware, networks, systems, and equipment that will make up the simulation can be built

Both conceptual validation and results validation have to correlate with system theory (the understanding of the subject being represented), but results validation also has to correspond to the available system data which were used in developing the system theory. Ideally system theory will correlate completely with all available system data. Sometimes system theory does not fully correlate with all available data. Then, results validation must compare simulation results with both system theory and system data.

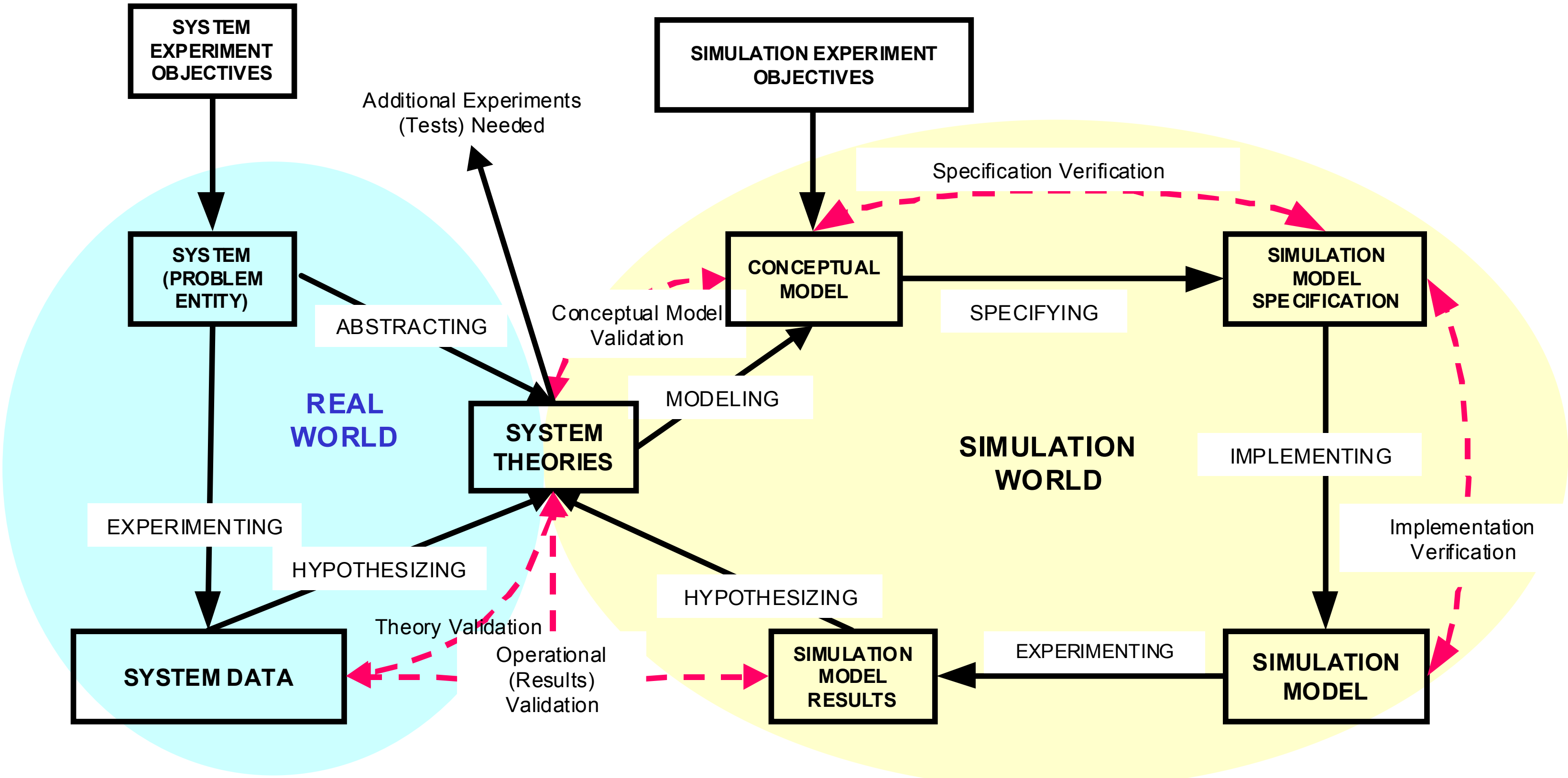
OBSTACLES TO QUANTIFICATION OF SIMULATION FIDELITY AND VALIDITY

Four *Validation Myths* which discourage efforts to quantify simulation fidelity and validity are identified here. These myths begin with the idea that any agreement between data and simulation results equates to validation (even in situations with sample sizes of one or a few, and data from situations that do not quite correspond to the simulated situation). The false conclusion drawn by many is that one does not need to strive to quantify simulation fidelity or validity since simulation results have "correlated" with real world data, no matter how spurious or aleatory that correlation may be. A second myth is the presumption that V&V is always too expensive, without any consideration of the possibility for a positive return on investment in V&V (even if large). This myth also discourages efforts to quantify simulation validity. The third myth is that V&V is meaningless because "political" factors, cost considerations, program importance, or other considerations are going to dominate the decision to accredit the M&S and guide use of its results. A fourth myth is that quantifying simulation validity is impossible because of the large numbers of simulation variables to be considered. We designate these obstacles as *myths* because they are not real roadblocks, though they are often-used excuses for not making serious efforts to quantify simulation fidelity and validity.

A more important obstacle to quantification of simulation fidelity and validity is lack of appreciation for the significance of quantifying simulation fidelity and validity. This limits investment in solving the problem, both efforts for a particular M&S application and in support for development of methods that facilitate quantifying simulation fidelity and validity. It is difficult to quantify simulation fidelity and validity, sometimes impossible to quantify simulation fidelity and validity to the degree desired. As better methods for quantifying computational accuracy and other components of M&S fidelity and validity are developed, e.g., [5], capability to quantify M&S fidelity and validity will increase
.

Figure 1. Real World & Simulation World Relationships

in Developing System Theories and Simulation Models with Verification and Validation (V&V)



Notes:

Experiment objectives should derive from validated requirements

Dotted red implies comparison, assessment, or evaluation

Validation is always relative to objectives/requirements/intended use

Diagram by Robert G. Sargent (Syracuse U) Jan 01

ENDURING ISSUES IN M&S FIDELITY AND VALIDITY QUANTIFICATION

Accepting Reality

It is often difficult for some to accept reality. When referent data are limited or have great uncertainties, capability to quantify M&S fidelity and validity will be severely restricted. Improving the M&S user interface or enhancing the impact of simulation results through sophisticated graphics and employment of virtual reality techniques cannot overcome validation limitations imposed by referent data lacks. Physics accepts the speed of light velocity limit; science fiction does not always constrain itself by such. Some in the M&S community seem more inclined to the science fiction approach than to the physics approach when it comes to recognizing M&S limitations.

Realistic Perspectives

Limited expenditure of VV&A resources can limit the level of confidence that can legitimately be placed in a simulation's correctness because the evidence to support greater credibility for simulation results has not been developed. Such limited credibility can increase the level of risk associated with using simulation's results to support system designs, operations based upon simulation results, etc. This perspective does not always prevail in the presence of strong interests promoting M&S acceptance.

Lack of a Compelling VV&A Business Case

It is difficult to educate program managers and other decision makers about the business case for VV&A so that appropriate timely VV&A decisions can be made because the M&S community has not developed a widely accepted model for how to account for VV&A-related costs (including generation of referent data) or how to determine return on investment for VV&A. For example, there is no standard method to compute return on investment for costs avoided by requirements V&V efforts that prevent potential faults by correcting M&S requirements before design and coding begin. Seat of the pants estimations suggest that such V&V efforts can result in significant reductions (half or more) of design and code re-work costs -- savings which probably would be at least an order of magnitude greater than the cost of the requirements V&V effort. Lack of a factual-based business case for VV&A is one reason that the myth that V&V is too expensive continues to be pervasive.

Personnel Competence.

Both M&S and its older, more mature relative, software engineering, are relatively immature as distinct disciplines. There are on-going debates within both disciplines about whether the discipline is mature and stable enough for initiation of professional certification for its practitioners. This lack of established and enforced professional practices, especially in M&S V&V, means that much of what is done is more representative of the performance of amateurs, some gifted and others not so talented, than the work of true professionals. An appalling result from the Military Operations Research Society (MORS) 1999 workshop on M&S V&V (SIMVAL 99) was evidence indicating limited knowledge by V&V practitioners of contemporary V&V technology and tools, and their lack of background to qualify them for effective use of formal methods and advanced techniques in V&V, [9]. This raises questions about competence of V&V personnel in use of statistical, logical, and other mathematical techniques for quantifying simulation validity.

REFERENCES

[1] Gross, D.C., et. al., "Report from the Fidelity Implementation Study Group," *1999 Spring Simulation Interoperability Workshop*, March 1999, Paper 167.

[2] Roza, Z. C., D. C. Gross, and S. Y. Harmon, "Report Out of the Fidelity Experimentation ISG," *2000 Spring Simulation Interoperability Workshop*, March 2000, Paper 151.

[3] Taylor, Barry N. and Chris E. Kuyat, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," National Institute for Standards and Technology (NIST) Technical Note 1297, 1994 Edition.

[4] Oberkampf, William L. et al, "Estimation of Total Uncertainty in Modeling and Simulation," Sandia National Laboratories Report SAND2000-0824, April 2000.

[5] Easterling, Robert G., "Measuring the Predictive Capability of Computational Models: Principles and Methods, Issues and Illustrations," Sandia National Laboratories Report SAND2001-0243, February 2001.

[6] American Institute of Aeronautics and Astronautics (AIAA) Guide G-077-1998, "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," January 1998.

[7] Trucano, Timothy G. et al, *Description of the Sandia Validation Metrics Project* (DRAFT), 30 April 2001.

[8] Jaynes, E. T., "Probability Theory as Logic," *Proceedings of the Ninth Annual Workshop on Maximum Entropy and Bayesian Methods, Dartmouth College, New Hampshire, August 14, 1989,* Paul F. Fougere (Ed.), Kluwer Academic Publishers, Dordrecht, Holland, 1990. Revised and extended version, May 1, 1994, available at http://bayes.wustl.edu/etj/articles/prob.as.logic.pdf.

[9] Pace, Dale K. and Priscilla Glasow, "SIMVAL 99 Final Report," Military Operations Research Society (MORS), March 26, 1999.