Introduction

Referent has a number of connotations, each of which may be appropriate in the context in which the term is used. The definition most often related to modeling and simulation (M&S) is:

Referent: A codified body of knowledge about a thing being simulated.

This is the definition for *referent* in the Modeling and Simulation Coordination Office glossary of M&S terms.[1] The definition derives from work on M&S fidelity in the Simulation Interoperability Workshop during the late-1990s.[2] In M&S, the referent has two functions. It supports M&S development (and modification), and it supports M&S assessment.

In supporting M&S development, the referent is the information upon which development of the M&S is based. That information determines the entities, processes, and interactions that should be included in the M&S to represent the simuland (the thing being simulated). That information determines the algorithms and data that should be used in the M&S. For clarity, it is helpful to call the referent used to support M&S development or modification the *development referent*.

In supporting M&S assessment, the referent is concerned with M&S fidelity, validation, and accreditation. The referent provides the information with which M&S results are compared to determine M&S fidelity and validity. That comparison provides factual information to support an accreditation decision. For clarity, when a referent is used in this way, it is helpful to call it an *assessment referent* or a *validation referent*.

This special topic is focused on referent use as a validation referent. This Special Topic is primarily for those who will be involved in DoD M&S validation and accreditation assessments. It provides specific guidance so that those who select, describe, approve, and use validation referents in M&S validation and accreditation assessments can do so effectively and efficiently. A check-list for validation referent identification and specification is provided at the end of the Special Topic.

An Expanded Definition

M&S validation is defined as "The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model."[3] This connotation for M&S validation is generally accepted, both within DoD and elsewhere, but sometimes people modify the connotation in various ways. For example, some restrict the term "validation" to situations in which M&S results are compared to high quality test data. Every such modification either adds to or subtracts from the connotation of M&S validation in the DoD definition and should be avoided when addressing M&S in DoD. Even those outside DoD would do well to use the term "validation" with the DoD connotation to avoid confusion because of the extensive use of the term with DoD M&S. If the situation is such that something different than what is meant by validation as defined by DoD is needed, then that something should be given a different label than validation. In the situation cited above, "data-based validation" is an appropriate way to describe the modification to the DoD definition. Doing such would avoid the confusion that arises when a different connotation is ascribed to a widely used term like validation.

The DoD definition for validation says what validation is, but does not describe how the accuracy of real world representation is determined. Determination of representational accuracy

requires comparison of the M&S, its data, and results with something. That something is the validation referent. The referent definition presented in the Introduction is not the best definition for the context of a referent used to support validation and accreditation assessment because the definition is too general. An expanded definition of validation referent will be more useful.

An expanded definition for *validation referent* shown below is based upon a 2004 study sponsored by the Defense Modeling and Simulation Office that involved participants from four nations.[4] U.S. participants came from Air Force, Army, and Navy organizations, Missile Defense Agency, National Academy of Sciences, academia, and industry.

Validation Referent: The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the M&S from the perspective of validation assessment for intended use of the M&S.

Various words and phrases from this definition are discussed below to ensure that the expanded definition for validation referent is fully understood. The words and phrases are in **bold font**.

The **information** used as the validation referent may consist of:

> "data" (observations of the simuland, either under controlled circumstances as in tests and experiments, or under natural or operational circumstances),
>
> theories as expressed in algorithms that describe characteristics, behaviors and relationships (preferably theories validated against observations of the simuland),
>
> simulation results (preferably from simulations that have been objectively and quantitatively validated by comparison with data about the simuland),
>
> expert human knowledge which in M&S assessment is described as subject matter expert (SME) estimations, or various combinations of these.

Consequently, the quality (accuracy, reliability, comprehensiveness, credibility, etc.) of the information in a validation referent may vary from one validation referent to another, and within a validation referent. In cases where explicit observations and theories do not provide a comprehensive and sufficiently reliable description of the reality represented in the M&S (i.e., the simuland), information from theory, other M&S, and SMEs may have to serve as the validation referent or part of it. Such SME information may not be explicitly articulated and systematically organized since it is a form of qualitative assessment, and the basis for qualitative assessments usually are not as explicitly described as in quantitative assessments.

M&S communities that deal mostly with physics-based models that may in principle be compared against quantitative data tend to be less willing to accept SME information and other qualitative assessments as part of the validation referent than are other M&S communities. The computational science and engineering community, which has major verification and validation (V&V) concerns in applications of computational fluid dynamics and computational solid mechanics simulations, is an example of that kind of M&S community. It should also be noted that the computational science and engineering community has been more prolific in V&V publications than most M&S communities. Publications of this community include V&V guides from professional communities as well as the only college/graduate-level textbook on V&V.[5, 6, 7]

However, subjective assessment enters even in the most quantitative situations. Results from essentially every test are judged for acceptability so that the results are not contaminated by bad data. Results from faulty instrumentation, from errors in installation of the instrumentation, etc.

are removed from the data considered as valid test results. Sometimes rationale for removal of such results may not be clear. At times, results have been removed merely because the results were not in the range expected by those conducting the test. The impact of this qualitative judgment on quantitative test results is seldom addressed explicitly.

The information used as a validation referent is a **codified body of information**. Codify has several connotations. It implies system and organization, both useful aspects of information to be used as a validation referent for M&S validation assessment. Codified often implies authority and is frequently associated with laws. Wherever appropriate and possible, the validation referent should be information drawn from credible authoritative data sources.

Unfortunately, a codified body of information many not cover the full spectrum of the intended use domain, even when the codified body of information draws upon theory and SME estimations as well as observations and test data. In addition, the codified body of information may have contradictions (or apparent contradictions) among different information items in the collection. Such aspects have to be addressed when encountered in validation referent identification, selection, and specification.

The definition above deals with the "**reality represented in the simulation**." This refers to the simuland. Typically that *reality* includes actors/systems/entities interacting with other actors/systems/entities by various processes through or in one or more environments. The validation referent pertains to all of these: actors/systems/entities, processes, interactions, and environments. Connotations associated with these may vary by M&S type and by the kind of application. For example, the M&S operator would not be part of the validation referent for a batch-run constructive M&S, but the M&S operator might need to be part of the validation referent in an interactive M&S (such as a game or war game).

Three challenges arise from the validation referent as the **best or most appropriate** information available that describes **characteristics and behavior** of the reality represented in the M&S **from the perspective of validation assessment for intended use of the simulation.**

1. Development and organization of the **best** information possible may be too costly and/or take too long for schedule and resource constraints of a particular M&S application. This leads to use of an *adequate validation referent*; one that is not the "best" information possible but one that is adequate as the basis for validation assessment within the context of the M&S intended use. Such an adequate validation referent can be the "most appropriate information available." Such an adequate validation referent may be the best information available within the time and resource constraints for a particular M&S application.

   The idea of an "adequate validation referent" increases the importance of exact and precise specification of the M&S intended use. A poorly specified intended use increases the risk of decision error, i.e., the risk either that an M&S will be judged acceptable (valid) for the intended use when it is in fact not acceptable, or that an acceptable M&S will be judge unacceptable. Consequences of such mistakes depend upon the impact of the M&S application. There may be little consequence from such a decision error for an M&S that only produces background information about a subject, but catastrophe could result from such a decision error about an M&S used as part of a real-time decision aid for a safety critical system. The ultimate authority for the acceptability of a validation referent is the accrediting activity (or Accreditation Authority), the person or organization

responsible for the decision that the M&S is appropriate for the intended use. This Special Topic develops rationale for identification, selection, and description of a validation referent, which also serves as a basis for its acceptance by the accrediting activity.

2. The **most appropriate** information has adequate fidelity to serve as the basis for comparison in validation assessment of the M&S for its intended use. It is a truism that one cannot demonstrate greater fidelity for M&S results than the fidelity of the validation referent to which those results are compared. Consequently, the validation referent must have greater fidelity (i.e., more closely represent the reality addressed by the M&S) than required for M&S results to be valid; otherwise the body of information selected is inadequate as a validation referent for intended M&S use. Vague specification of M&S intended use can make it difficult to determine required fidelity for an adequate validation referent. This leads to increased decision risk and makes it more likely that an inadequate M&S will be considered acceptable for the intended use. For more information see *Advanced Topics>Special Topics>Fidelity*.

3. *As a general rule, the most appropriate collection of information to serve as the validation referent is the least expensive set of information that has adequate fidelity to support M&S intended use.* Otherwise, the validation referent becomes gold plated and its development wastes resources. For example, if M&S results are to provide ballpark estimates of performance as background information, SME judgment may be an acceptable validation referent for such "back of the envelope" fidelity. Spending time and effort collecting data and manipulating the data into an acceptable format to serve as a validation referent for such an M&S could be a waste of resources if the SME judgment could provide an adequate assessment of M&S fidelity for its intended use.

Because the validation referent **describes characteristics and behavior of the reality represented in the M&S,** specification of the validation referent is particularly difficult for those M&S which generate new knowledge or additional knowledge about the reality represented in the M&S (vice the M&S simply being a reliable representation of that reality). This could be the case with an M&S demonstrating emergent behavior. This also could be the case if the M&S employs some form of judgment in uncertain conditions (such as representation of human decision making) or if the M&S portrays a future reality that is difficult to predict (such as social structure, military posture, etc.). There are two implications here for validation referent specification. One is that iteration of the validation referent specification may be necessary as new knowledge is generated. The hazards of such should be obvious. The other is that intended use of M&S with such uncertainties in the reality represented implies that the validation referent specification will have low fidelity. It is not the length of time projected into the future that is the issue; astrophysical simulations project conditions billions of years in the future with high fidelity. It is the uncertainty in the representation that determines the fidelity needed in the validation referent specification.

Additional Comments Related to the Expanded Definition

The expanded definition for validation referent presented above can apply to all varieties of M&S and their full range of potential applications, but how one identifies the information appropriate for a validation referent and selects an appropriate set of information as well as how

one describes the validation referent may vary with M&S type and application. Subsequent sections will discuss these topics.

This Special Topic focuses on the referent in validation and accreditation assessments. Consequently, how information in a development referent may be used to develop a model or simulation is identified or described is not addressed extensively. Obviously the development referent is related to the validation referent, but dealing with the development referent in detail is beyond the scope of this Special Topic. Nor is how M&S inputs should be selected addressed in this Special Topic, whether the inputs are hardwired in or are fed into the M&S as needed to run it. The focus on validation referent use in validation and accreditation assessment makes it necessary to consider issues related to statistical independence (or dependence) of information used for M&S development relative to information used as the referent in validation and accreditation assessments when M&S results are to be used to predict performance or behavior.

Many M&S communities, especially those in which models often have heuristic factors that are adjusted to make M&S results fit experimental data better (this process may be called "model calibration" as it is the case, for example, in computational solid mechanics) stress the importance of separating data used in such model calibration from the data used in M&S validation assessments.[6] The problem is similar to that encountered in clinical trials in the health field, in which randomization of patient involvement in a clinical trial (with its associated unpredictability) is used to "protect against the unpredictability of the extent of bias in the results of non-randomized clinical trials".[8] Many do not appreciate the issues associated with using the same information for M&S validation assessment as was used for M&S development or the limitations that such places on what can be determined about M&S accuracy or predictive capability under those conditions.

For editorial simplicity, validation assessment may be used to encompass all referent aspects of both validation and accreditation assessment. It is important to appreciate the difference between validation assessments and accreditation assessments. The validation assessment determines whether the M&S has adequate fidelity to support intended uses, and perhaps can even quantify the likelihood that the M&S has the required fidelity with a statement like "we have 90% confidence that M&S results will differ from the validation referent by less than 5% for any conditions within the specified application region." Such a quantitative statement is the most precise kind of validation assessment, and depends upon a very well defined intended use, an explicitly well defined validation referent, and a precise measurement capability for the M&S. Few validation assessments have had such quantitative precision.

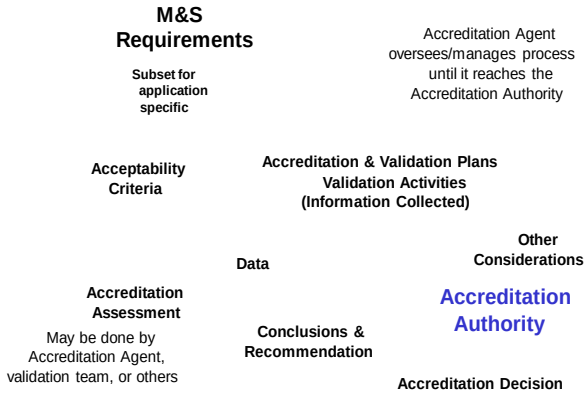
Accreditation assessment uses the validation assessment, but supplements it with risk, programmatic and other considerations to determine if the M&S is appropriate for use in a particular situation. *Whenever accreditation assessment is mentioned in this report, the only concern will be its validation assessment component – other aspects of the accreditation assessment are outside the scope of this Special Topic.* This makes it convenient to use the term "validation assessment" to refer to either or both validation and accreditation assessments. It is important to remember that validation assessment is always in the context of intended M&S use, and that the accreditation activity (or Accreditation Authority) is the one who determines the acceptability criteria for M&S intended use and, thus, in essence is also the one with final approval authority for the validation referent.

Validation Referent Identification and Selection

This section is concerned with three topics. First how to identify possible validation referents for M&S validation assessment; second, how to select an appropriate validation referent when there is more than one possible referent; and third, what to do when an appropriate validation referent does not exist or cannot be obtained within available time and resources.

Validation Referent Identification

Two factors make identification of M&S validation referent non-trivial: a) M&S requirements and b) variations in characteristics of the reality to be represented in the M&S (i.e., the simuland). Requirements identify what is to be represented in the M&S (actors, systems, entities, processes, interactions, environments, etc.) and the M&S purpose (which has implications for fidelity of representations and how the representations can be manipulated). A particular application may be driven by acceptability criteria (which are based upon M&S requirements). Figure 1 illustrates how these factors relate to one another in an accreditation assessment.



Accreditation Process in a Nutshell

Validation and accreditation assessments may be done in a formal process, as is often the case with a major M&S application, or they may be done very informally as often happens in a minor M&S application. In either case, the same kinds of processes are involved in deciding that an M&S is appropriate for a particular application.

To illustrate how these factors make validation referent identification non-trivial, consider the validation referent of the natural environment for an M&S whose purpose is to predict sensor performance capability. If a radar sensor is involved, natural environmental factors that could be pertinent include terrain (to determine where radar signals would be masked by terrain features), precipitation intensity, extent, and duration (since radar signals can be attenuated by precipitation), temperature and pressure gradients with altitude (these affect radar ducting), etc. These factors can depend upon geographic location and vary with time of day and of the year. They may change from one day/year to the next. For some purposes, a simple free-space radar range computation will be adequate (a computation which is not affected by any of the factors mentioned above); and, in other cases, all of these factors must be considered. In some cases, parameter values from standard handbooks or catalogues will provide the information needed. In

other cases, detailed observations must be made at the specific site where the radar will be located. In some real-time applications, the information may have to be current, not historical (neither recent history from the previous day nor history from years before will be adequate).

M&S purpose and the nature of the reality to be represented specified in the requirements determine what characteristics the validation referent should have (what parameters should be represented and how). Identification of possible validation referents seeks to satisfy these. Information sources which contain all pertinent factors for conditions specified provide an initial set of possible validation referents. M&S fidelity requirements then determine accuracy and uncertainty characteristics of the information needed for it to qualify as a possible validation referent.

When some information items are not available in a possible validation referent (or information items do not have needed fidelity, or information fidelity is uncertain), it will be necessary either to institute a test program to develop the needed information or to modify M&S requirements so that they can be satisfied by the available information. Such a modification in requirements may require the M&S to address only part of the problem instead of the whole problem, or the modification can take the form of allowing a theoretical value or SME estimation to be used for some parameters instead of requiring test data of a specified fidelity. For complex M&S, the validation referent is often drawn from a collection of information sources, not from a single information source.

Correlating the validation referent with M&S requirements is essential; otherwise, the M&S may not be able to support its intended use. Whether theory or SME estimation has adequate fidelity for the validation referent, or extensive use of precision data from tests are required for the validation referent to have adequate fidelity, is determined by the requirements. The validation referent must be able to support M&S intended use.

Validation Referent Selection

Selection of the validation referent is usually done on the basis of direction, convenience, economics, a decision to use proxy information for the validation referent, extent of coverage of the intended use domain, community acceptance of the information (i.e., its credibility), or some combination of these. Each of these is discussed below. Any possible appropriate validation referent must be acceptable in terms of scope, reliability, credibility and fidelity.

The possibility of multiple acceptable validation referents will be illustrated by considering possible validation referents for the length of a meter. From 1889 to 1960, the official international validation referent for the length of a meter was an international prototype of a meter with an estimated accuracy of less than 10^{-6} meter error. From 1960 to 1983, the international referent for the length of a meter was based up the wave length of ^{86}Kr with an estimated accuracy of less than 10^{-8} meter error. Since 1983, the international referent for the length of a meter has been based upon the speed of light in a vacuum with an estimated accuracy of less than 10^{-9} meter error. More often, people use a less accurate referent for measurements than the international standard. Such less accurate referents range from the ancient informal standards of the distance from finger tip to nose or a long stride to contemporary approaches such as a standard measuring tape, a yard stick, or a more precise measuring device. Depending upon the application, measuring devices (referents) with accuracies measured in terms of a few percent (or more) could be acceptable. One need not go to the expense of measurement accuracies associated with international standards for length if the measurements are only going

to be used to estimate how much paint will be needed to paint a wall. Measurement accuracies of 80-90% will be more than adequate for that application. This simple example illustrates how a variety of referents may be adequate for a particular application. Selection of the validation referent that is most appropriate would depend upon the circumstances; especially the M&S intended use.

Validation Referent Selection by Direction

When the validation referent selection is based upon direction, the validation referent selected often is specified by the M&S sponsor or user and approved by the accreditation activity/authority. As long as the validation referent specified in such direction has acceptable fidelity, there are no problems with validation referent specification by direction. Often a particular validation referent will be selected by direction for compatibility with other aspects of a large program or for compatibility with related programs.

However, potential problems can occur when a validation referent specified by direction does not have adequate fidelity to support intended M&S use. When this kind of situation is discovered, it is necessary to communicate with the one providing direction about the validation referent and explain the problem. It will be necessary to find an adequate validation referent, or to change the requirements so that the validation referent selected is adequate.

Inappropriate validation referent specifications by direction can be a problem when SMEs have to serve as the validation referent or to supplement facts and theories as part of the validation referent. Use of an inappropriate SME may be directed. If such direction is proposed, those responsible for validation assessment should inform the authority of concern about the direction and request a change to the direction.

Sometimes a SME may seem to be inappropriate when the SME in fact is appropriate. If only technical expertise is considered in judging SME qualifications, a particular SME might seem inappropriate. However, that SME might be appropriate because SMEs can be qualified by organizational or vested interest association as well as by technical expertise.[9] For more information see *Advanced Topics>Special Topics>Subject Matter Experts and VV&A*.

Validation Referent Selection by Convenience

When the validation referent is selected for convenience, the validation referent may be the one easiest to access, one already available vice one coming in the future, or one that the M&S team knows and is comfortable working with. As long as the validation referent has acceptable fidelity, no issues arise from selecting the validation referent on the basis of convenience.

Validation Referent Selection by Economics

Since cost of information is usually related to its quality (better costs more), it makes economic sense to select the least costly information that satisfies validation referent content and quality requirements (even if such is not the best quality information in some sense). The key is that the information is adequate in scope and accuracy to serve as validation referent for M&S intended use. As long as a collection of information satisfies the adequacy test for M&S intended use, it is an appropriate validation referent. The most appropriate validation referent typically will be the least expensive one. If economic considerations preclude use of an adequate validation referent, then M&S requirements or intended use may have to be modified before an appropriate validation referent will be viable.

Validation Referent Selection by Proxy

Very often, it is not possible to identify a validation referent for the specific performance under study. For example, this will nearly always be the case when the reality represented by the M&S is a future system. In that case, performances of similar tasks on existing systems may be adapted to provide a referent. As an example, when assessing reaction and decision times for operators of a new combat system or machinery control system, there may not be any data regarding operator performance; however, past experience with operator performance in the accomplishment of similar tasks on existing systems may be used as the validation referent for simulations of the new system.

Validation Referent Selection by Coverage Extent

For situations in which the M&S intended use application domain is very large, it may be very desirable to have as few information sources for the validation referent as possible in order to minimize problems of creating coherent information for the validation referent. Determination of information adequacy when the information covers a broad spectrum of parameter values always needs careful attention. Often the fidelity of information varies with parameter value. Thus, coverage extent has to be within the context of acceptable fidelity.

Validation Referent Selection by Community Acceptance

In some situation, community attitude toward M&S results is more important than in other situations. In situations in which community attitude is important, selecting the acceptable validation referent with the greatest community acceptance will facilitate community acceptance of M&S results since they will trust the standard for validation assessment. Confidence in the assessment standard facilitates confidence in M&S results.

Validation Referent Selection by a Combination of Considerations

More often than not, validation referent selection will involve more than one of the factors mentioned above. There is no general rule for how to weight the different factors. That will depend upon the particular situations. The over-riding fundamental principle is, regardless of what factor or factors are involved in referent selection, any validation referent selected must be adequate. Only adequate validation referents should be used. If it requires modification of M&S intended use and requirements for available validation referents to become adequate, then that has to be done.

A Caveat Regarding Validation Referent Fidelity

Any useful validation referent has to have adequate fidelity if it is to be a meaningful standard of comparison for use with M&S results. It is easy to think of fidelity only in terms of parameter or characteristic accuracy. Such considerations impact the kinds of information sources that can produce information with acceptable levels of accuracy and credibility. This can be a problem when the only information sources are ones with questionable accuracy and credibility, sources such as SME estimations.

However, there are other aspects of fidelity that are of concern for referents beyond information accuracy and credibility, such as scope, resolution, and context.

- **Scope** is concerned with the range of parameters or applications of concern for the referent. For example, information about a material, such as H_2O, varies with material

phase (solid, liquid, vapor/gas), which is a function of temperature, pressure, salinity, etc. The scope of a validation referent for H_2O would indicate the range of relevant physical conditions over which it would be applicable. Discussion above of selection of the validation referent by coverage extent focused on coverage at the individual parameter level. Scope can involve multiple dimensions.

- **Resolution** is concerned with the level at which distinctions can be made in the information of the validation referent. Some information might describe an item's behavior and characteristics at atomic levels; other information at the component level; and other information may be at subsystem, system, or unit levels. In dealing with M&S representing human behavior, the referent might have information about individual behavior, or only about team or group behavior. Resolution of information in the validation referent must go to the lowest level for which M&S validation is being assessed.
- **Context** addresses the environment within which the validation referent information is applicable. Sometimes the context is simply a set of assumptions; sometimes the context is a physical condition (such as pressure). The difference between context and scope is simple. Context is not specified in the variables used in the description of actors, systems, entities, processes and interactions described by the information in the validation referent, but context is concerned with parameters and factors that might influence those descriptions. For example, the information about how people perform certain tasks that is used as a referent for M&S description of human behavior representation probably varies with whether those people were observed by their bosses when their performance was measured, whether the people felt their performance was critical (such as a promotion depended upon it), or other such conditions. Such conditions are part of the context for that information.

Consequences Arising From Lack of an Adequate Referent

Sometimes the information that is available or likely to become available within expected time and resources before M&S intended use will not be adequate to support validation assessment for M&S intended use. In such a situation, one of the following three options must be taken. It is recognized that none of the options is a desirable choice, but they are the only ones available. An attempt at validation assessment without an adequate validation referent will be capricious, regardless of how it may be described.

- Accept that validation assessment will not be possible because of lack of an adequate validation referent.
- Provide additional time and resources so that an adequate validation referent can be developed.
- Modify M&S intended use (and associated requirements) so that available information can provide an adequate validation referent.

The Accreditation Agent, V&V Agent, or others involved in identification, selection, and preparation of the validation referent must clearly explain the situation so that the appropriate authority can decide how to deal with the situation.

Validation Referent Specification and Description

This section addresses a number of topics:

- Where in the M&S lifecycle the validation referent should be specified.
- What should be included in description of the validation referent.
- How description of the validation referent depends upon M&S variety, M&S application, and availability of referent data.

Sometimes description of the validation referent will be pointers to detailed information elsewhere (such as test documentation and results), with the referent description in M&S documentation merely providing the pointers and explaining how the information that is indicated fits in the validation referent. Other times, M&S documentation of the validation referent will include both the information and its description. Discussion in this Special Topic applies to both situations.

Validation referent descriptions have five fundamental aspects: 1) context, 2) domain coverage, 3) attribute distinctiveness, 4) parameter uncertainty quantification, and 5) information coherence. Specification or description of the validation referent should include all five aspects. Each of these aspects is discussed below.

Some general characteristics should apply to every validation referent specification. Identification and specification of a validation referent should be definite and unambiguous. Just doing this consistently will bring improvement over what is often done in current M&S practice. If information in the validation referent only uses part of information from a particular source (such as results from one test in a set of tests), what is not to be used should be clearly and specifically delineated. It usually helps to add an explanation of why this is being done. For example, if the flight profile of a missile is to come from a particular test, but the radar signatures of the missile in that test are not appropriate as part of the validation referent for the missile represented in the M&S (perhaps the signatures of the missile in the test were augmented to facilitate range tracking of the missile during the test), that should be noted explicitly and clearly in the validation referent identification and description. This is to eliminate any possible confusion about what is to be used as the validation referent.

1. Context

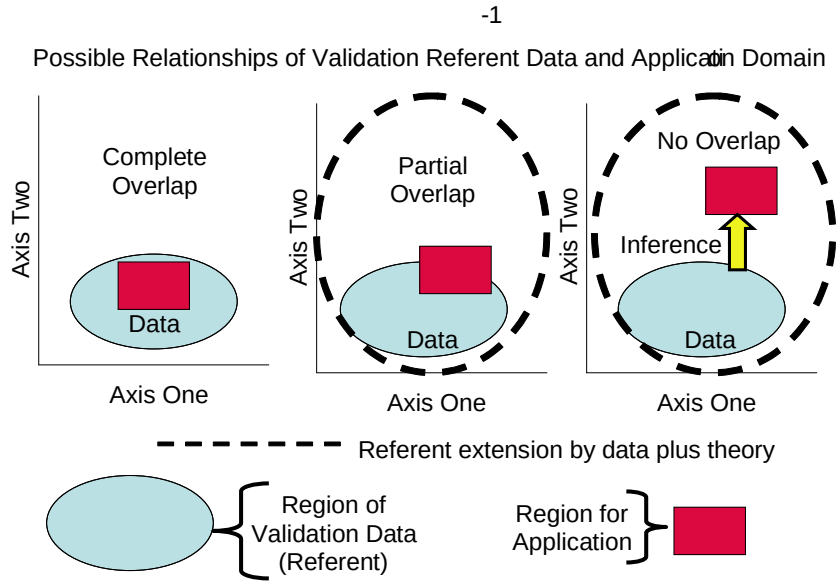
It is always important for a validation referent to specify its context. This is particularly important for situations in which there may be debate or uncertainty about what should be the validation referent. The context will range from information about conditions under which human performance information is collected to physical conditions (such as temperature, pressure, radiation, etc.) that might influence measured parameters in the validation referent which are not addressed specifically by M&S algorithms. However, care must be taken to avoid specifying the validation referent too narrowly.

2. Domain Coverage

An M&S has a specified application domain. Its intended use determines the appropriate application domain. The application domain is always multi-dimensional. Validation referent description should indicate what portion of that application domain is addressed by the referent. For example, test data used as a validation referent may reflect steady state, smooth, undisturbed

flow conditions for a parameter (such as fluid volume passing through a pipe, traffic on a road, time delay waiting for a technician in on-line support, etc.) but such data are not appropriate for use as a referent in transition, turbulence, disturbed flow conditions (such as might be experienced if an obstacle were in the pipe or on the road, or the on-line support shift is short handed).

Limitations in validation referent domain coverage can force consideration of inference. The figure below illustrates three possible relationships between data for a referent (the blue ovals within the solid lines) and intended use (red boxes) in the application domain. If the validation referent is restricted to data, only when the data completely overlap the application area can there be high confidence in a quantitative assessment of the relationship between simulation results and the validation referent. In the other two situations (partial overlap and no overlap), indeterminate uncertainty is present for applications outside the data region. If reliable theory exists, then either of these cases might be capable of being reduced to the equivalent of the complete overlap case by use of theory for any point outside the data region.



Data Application Domain Overlap Possibilities

General parameter regions are fairly easy to understand. For example, if a simulation application is to address what happens when bodies collide (as with a missile defense kill vehicle hitting its target), the range of interest for collusion velocities may vary from nearly zero (as could happen in a scenario in which the kill vehicle approaches from behind the target) to very fast (10 km/s or more for a fast interceptor against a ballistic missile in a head-on encounter). Data from full-scale tests may be very limited, not only in the number of tests, but also in the portion of the speed regime for which there are tests and in the availability of precise information from the tests. Data from surrogate tests (such as sled-tests or light gas gun experiments) may supplement the full-scale test data, but they introduce uncertainties into the data because of test artifacts (such as need to scale results, differences between the surrogate and the real object, etc.) and such tests may not fully cover the parameter regime of interest. Then theory or perhaps very high fidelity M&S results may be used for “data” for parts of the parameter regime that testing

(either full-scale or surrogate) cannot address. And finally, expert opinion may be used to fill in any remaining information gaps in the domain (and to reconcile any discrepancies among the various kinds of information mentioned).

Description of the referent will indicate how information about the M&S actors/systems/entities, processes, and environments throughout domains of interest are addressed. This description also identifies the information sources to be used for each portion of the domain, gaps that may exist, methods to reconcile information from different sources, and methods to combine information in the validation referent.

Whether a partial or no overlap situation is adequate as validation referent (supplemented by theory and SME estimations) depends upon M&S intended use and the level of fidelity required for validation assessment. An inference assessment will be theoretical. Intended use must be compatible with that if the validation referent involves inference in any way.

3. Attribute Distinctiveness

The validation referent is concerned with actors/systems/entities, their interactions, processes, and environment(s) of the reality represented in the M&S. As illustrated earlier, not every possible attribute of these is significant for the validation referent. All attributes needed to satisfy M&S requirements should be specified for the validation referent, and described fully for the domain coverage required with all pertinent variations indicated. For example, if the object size or color changes with temperature, etc. and the object's size or color is important for intended M&S use, that attribute trait should be specifically noted.

4. Parameter Uncertainty Quantification

There are two uncertainty dimensions in fidelity and validation assessments. One dimension is M&S uncertainty. These uncertainties arise from imperfect algorithms, computation characteristics (such as table look-up errors), input errors, etc. Often when M&S results are compared with a standard (theoretical curve, test data, etc.), it is assumed that all error or uncertainty is a result of M&S uncertainties. This usually is not the case. The other uncertainty dimension in fidelity and validation assessment is uncertainty in the referent. Only when referent uncertainties are characterized can M&S uncertainties be characterized. When both uncertainty dimensions are fully characterized, then fidelity and validation assessments can be performed rigorously and fully characterized.

A great deal of attention has been given to quantifying uncertainties, both M&S uncertainties and referent uncertainties. Various validation metrics have been suggested, guidance in regard to validation experiments has been developed so that referent information with characterized uncertainty may be known, etc. A proper validation referent description always includes specification of parameter uncertainties in validation referent information, even when description of parameter uncertainties is limited (such as simply bounding the uncertainties, e.g., the uncertainty is expected to be less than x %). Sometimes the available information does not permit a better characterization of the uncertainty.

Sometimes validation referent uncertainty is characterized by the kind of information used in the referent. Three basic kinds of information are: 1) quantitative test and experiment data, 2) fully formulated theory that has been correlated with substantial evidence and data, and 3) and loosely formulated theory including SME estimations.

Standard statistical processes can be applied to quantitative test experiment data to estimate the uncertainties (and accuracies) of a validation referent based upon such information. The more data one has, the better the estimation of such uncertainty and accuracy.

Standard statistical processes can be applied also to fully formulated theories to estimate the uncertainties (and accuracies) of a validation referent based upon such information. A fully formulated theory has demonstrated its reliability by being successfully applied in numerous circumstances by a wide variety of people.

Loosely formulated theories and SME estimations often have limited repeatability, i.e., different people get different results when these assessment methods are applied and the uncertainties (accuracies) of these assessment methods may be difficult to estimate. Uncertainties are less when SMEs estimates are employed in a validation referent if guidance for proper SME usage, are followed. Likewise, multiple SMEs can be used, which permits statistical analysis of SME estimations that provides some indication of variability in the SME estimations and reveals trends in SME estimations. For more information on SME usage see *Advanced Topics>Special Topics>Subject Matter Experts and VV&A*.

5. Information Coherence

As indicated previously, information for the validation referent may come from multiple sources. Some will be redundant (same parameters for the same part of the same domain), some will be supplementary (different parameters for the same part of the same domain, or same parameters for different parts of the same domain), and some will be disjoint (different parameters and different domains). Some information will have parameter uncertainties quantified, and other information will not. Information coherence is concerned with how information is combined so that information about a particular aspect of the validation referent (parameter, actor, system, entity, process, interaction, environment, etc.) makes sense and is compatible with other information in the validation referent. The validation reference description should explain how information combination is done so that information coherence is achieved.

The two examples below illustrate some of the ways this might be done. Neither example tries to show a best or preferred way for combining information. Decisions about what is best or most appropriate in a particular situation will depend upon details of the situation.

Example 1. Information about parameter x will be taken from three tests, all of which address how the parameter varies with respect to y over the same set of values for y. In Test 1, uncertainties in the measurement of x are fully characterized. Tests 2 & 3, which were done at the same test facilities using the same equipment and test personnel, do not characterize x uncertainties nor describe environmental conditions fully. Options for referent information about x include:

Use all data from Tests 1-3 and ignore uncertainties about x

Use only data from Test 1 and fully characterize x uncertainties

Use a plot of x from Test 1 with uncertainties indicated, overlaid with data from Tests 2 & 3, as a basis for an equation to characterize x vs y

Example 2. Information about parameter x is available from Test 1 for a portion of the domain of interest. Information about a parameter y, to which it is believed that x is proportional, is available for a portion of the domain of interest not covered by Test 1 (but without any

information for the part of the domain that is addressed in Test 1). Two SMEs gave widely divergent opinions about what multiplication factor should be assumed in the proportional relationship between x and y. Options for referent information about x include:

Use only data from Test 1

Use both data from Test 1 and a band for x determined by proportionality to y (with the two SME estimations establishing the edges of the band)

Declare that x is unknown

A proper description of the validation referent will explain how information from different sources is to be combined coherently. Obviously general guidance for how information should be combined is beyond the scope of this special topic; but it is clear that explicit description of how such information will be combined is an important part of a validation referent's description.

When and Where to Specify the Validation Referent

There are many ways to describe the M&S life cycle. The M&S lifecycle can be broken into eight phases, which may be passed through serially or repetitively. More than one of the phases may be occurring concurrently. These phases occur regardless of M&S development paradigm employed, whether a serial paradigm or an iterative one. The phases are:

- Requirements (expression of needed or desired M&S capabilities)
- Planning
- Conceptual Model
- Design
- Implementation
- Testing
- Use
- Modification

Early specification of the validation referent in a new M&S development (or in a major modification of an M&S) is needed in order to support conceptual validation assessment of the M&S. The validation referent should be the same for conceptual validation assessment as it is for results validation assessment and accreditation assessment. Otherwise the standard by which the M&S is evaluated changes from one aspect of assessment to another. That can have negative impacts on efficiency of M&S development as well as reducing M&S credibility.

For a legacy M&S, early specification of the validation referent means that the validation referent should be specified early in accreditation planning. Since by definition, a legacy M&S already exists, specification of the validation referent will be after the M&S has been developed.

Delay of validation referent specification until late in the M&S development for a new M&S development or modification (or until late in accreditation planning or execution for a legacy M&S) can introduce a variety of problems. For a new M&S development or an M&S modification, it will mean that conceptual validation will be performed using a standard (referent) that may not be the same as the standard (validation referent) to be used in assessing

the completed M&S. Opportunities for unnecessary problems and extensive rework abound in such conditions. It is much better to avoid these potential problems by identifying and specifying the validation referent early in the M&S life cycle.

For both legacy and new/modified M&S, late specification of the validation referent can delay M&S accreditation because needed information which could have been obtained by appropriate early planning is not available. Such delay in validation referent specification also can increase VV&A costs by not facilitating efficient V&V planning and execution.

Early specification of the validation referent for new M&S developments (and for M&S modification) normally occurs during requirements or planning phases. Typically specification of the validation referent will occur in conjunctions with development of acceptability criteria if the validation referent has not been specified earlier. It is also helpful to specify the validation referent in documentation of the conceptual model since this will ensure an appropriate focus and context for M&S conceptual validation since that should be performed relative to the same standard as results validation later.

Where the validation referent is specified may vary. In some cases, the validation referent will be documented by itself with pointers to it from the accreditation plan, V&V plan, and conceptual model. The DoD standard for documentation of VV&A does not use the term "referent," but its templates for the accreditation plan, V&V plan, V&V report, and accreditation report all include "Basis of Comparison" as Appendix C.[10] This appendix is where the validation referent would be described in these documents. In other cases, the validation referent may be documented in one of three places with pointers to it in the other two (the three places are the accreditation plan, the V&V plan, and the conceptual model). And yet in other cases, the validation referent may be documented in each of the three places indicated (accreditation plan, V&V plan, and conceptual model).

For accreditation of a legacy M&S, the validation referent should be specified in the accreditation plan. Depending upon the circumstances, the validation referent may also be described elsewhere, or at least there may be other pointers to the validation referent described in the accreditation plan.

Validation Referent Specification Dependence on M&S Type and Application

The spectrum of M&S use within DoD is vast. Every M&S technology is used somewhere within DoD M&S: from desktop and hand-held devices to super computers, from single computer M&S to large Live-Virtual-Constructive exercises involving many M&S and live forces, etc. Applications of DoD M&S include warfighting M&S (wargames, M&S supported decision aids, M&S to explore tactics and strategy, etc.), M&S involvement in all aspects of science, engineering, manufacturing, and logistics, M&S in education and training (from flight simulators to remote learning M&S), business M&S (planning, finance, personnel, etc.), and more. Guidance in this Special Topic about validation referents has to be applicable to the whole spectrum of DoD M&S. This impacts how specific that guidance can be since some guidance might apply only to certain kinds of M&S, and therefore could not be included in the guidance of this Special Topic.

It has been suggested that validation referent guidance depends upon the kind of M&S and upon the kind of application to which the M&S is applied. Five varieties of M&S and five kinds of

applications were selected to demonstrate appropriateness of the guidance that has been presented. Insights gained from considering validation referent guidance for these ten situations were incorporated into the guidance presented earlier in this Special Topic. The variety of these ten gives confidence that the guidance about validation referents in this Special Topic can be applied for any M&S or M&S application with DoD.

The five varieties of M&S considered were:

1. M&S which make extensive use of adaptive programming
2. M&S which involved human behavior representation
3. Distributed M&S
4. M&S which make extensive use of aggregation
5. M&S with system/hardware/software/people in the loop

The five areas of M&S application considered were:

1. Computational science and engineering applications
2. Engineering level applications
3. Gaming & training applications
4. Military theater-level and campaign-level applications
5. Non-physical applications (such as economic M&S)

Each of these ten areas is discussed briefly and with special emphasis on points related to their validation referents.

Validation Reference Dependence on M&S Variety

Information needed for the validation referent can vary with the kind of M&S. This is illustrated by discussion of validation referents for five varieties of M&S. As illustrated by discussion of the five kinds of M&S, guidance about validation referent identification, selection, and specification/description presented earlier applies in all cases.

M&S Which Make Extensive Use of Adaptive Programming

There are many adaptive programming techniques employed in modern M&S: artificial intelligence, knowledge-based systems, genetic algorithms, fuzzy logic, agent-based simulation, neural nets, etc. Sometimes such an M&S will be labeled as a complex adaptive simulation. Some of these approaches learn. Some evolve, and change the M&S structure and algorithms. Some M&S with adaptive programming are designed to demonstrate how things behave where their behavior is unknown.

At present, there is no standard kind of information that is considered the validation referent for such M&S. It depends upon the specific use to which the M&S is being applied.

Sometimes the validation referent is a description of low level/detailed part behavior or characteristics. This can give M&S User confidence that the M&S replicates such behavior or characteristic properly. Then the M&S can be used with some level of confidence to explore

what might happen at aggregated levels of behavior under various scenarios as the M&S evolves through them even when there may not be a specific referent for behavior at the aggregated level.

Sometimes aggregate level information will be selected as the validation referent. Then many runs are made with the M&S and the one that most closely matches the validation referent is selected and examined to discover the path that led from initial conditions to that result. This approach has provided valuable insights about biological processes.

These two examples show how the kind of validation referent information can vary with M&S intended use. However the validation referent information may be selected, it should be specifically identified and thoroughly described in accordance with validation referent guidance presented earlier That guidance works for this kind of M&S in the same way it works for other kinds of M&S. The role planned for the validation referent in comparison with M&S results should be specified in detail.

M&S Which Make Extensive Use of Human Behavior Representation

Validation referents are difficult to identify for human behavior representation in M&S, as indicated by the conclusion of a recent report: "There is currently no common way to implement, measure and validate human behavior in models." Conclusions about human behavior representation validity in this report also note, "Current constructive entities are often brittle, showing unrealistic behavior for even slight departures from the design space due to an over-reliance on simple, rigid rule sets and strict behavioral templates that capture standard operating procedures and idealized, purely doctrinal behavior. Validity on the constructive and predictive levels is often weak because of the prohibitive cost of extensive validation and because of limitations in experimentation with threatening conditions."[11]

The difficulties in deciding what information to use for comparison with M&S performance show the importance of the validation referent specification explicitly addressing both the context of the information and its scope (which parts of the application domain of M&S intended use is covered by the information) – this illustrates the value of the five aspects of validation referent description presented earlier. Careful description of validation referent context and scope helps to ensure that an appropriate perspective is used when M&S results are assessed relative to the referent.

Validation referents described in accordance with guidance presented earlier will help the M&S User and Accreditation Authority think through what the M&S is to be accredited for. Is M&S human behavior representation to be validated at the level of individual behavior (such as the action of an individual soldier or crew member)? Or is the M&S to represent the behavior of the group (such as the crew of a combat vehicle)? Often there is more reliable information about the behavior of a group than of an individual. Clarity in such matters can help to determine what information is pertinent and appropriate as a possible validation referent.

Even though human variability and limited awareness of context factors that can impact human behavior create much uncertainty in human behavior representation validation referents, the guidance provided in this Special Topic and validation referent specification and description apply to human behavior representation M&S as well as to other kinds of M&S.

Distributed M&S

The validation referent is the norm against which the results of M&S are compared for the purposes of validation. This subsection discusses general principles involved in identifying and selecting an appropriate validation referent for a distributed M&S so that it will be clear that guidance presented earlier about validation referent identification and specification also applies to distributed M&S. Specific guidance may be available for a distributed M&S employing a particular paradigm. Such guidance for High Level Architecture distributed M&S may be found in the VV&A Overlay for the Federation Development and Execution Process.[12] The Federation Referent is used in conjunction with development of acceptability criteria, in validation of the Federation Conceptual Model, and in validation of federation results. The material provides general guidance similar to that presented in this Special Topic, but does not provide detailed guidance about information that should be contained in the Federation Referent or its format. Ideas presented in this Special Topic and those about the Federation Referent are fully compatible. The Federation Referent has a special challenge in regard to combining information from different sources for the validation referent, i.e., how to combine information from validation referents for the federates that comprise the federation. The importance of credibility of the Federation Referent with the User/Sponsor of the federation is emphasized.

Validation Fundamentals for Distributed M&S

Distributed M&S pose a special validation assessment problem in that they are intrinsically more complex than many unitary M&S because they involve both the interaction of multiple processes and the infrastructure which makes the interaction possible. A coarse assessment of M&S performance may be obtained if the results can be compared to a validation referent consisting of a previous "real life" event, either from an operational event or in a test/training situation. Because of limited controls in real life events, there is significant variability in data from "live" events. The intrinsic variability of real life events places limitations on the ability of one or a few real life events to serve as a referent. The variability must be addresses.

Problem decomposition is one of the basic procedures employed in scientific methods. It can also be applied to the validation referent for a distributed M&S. It is a particular application of what some call "piecemeal validation." Simpson's Paradox is a potential problem for any piecemeal validation, but often it is not considered. Simpson's Paradox is the potential problem of the direction of a conclusion from examining pieces being at odds with the direction of a conclusion for the whole.[13] It is always a concern for piecemeal validation and should also be a concern in the decomposition approach to the validation referent for distributed M&S.

Because of the variability of real life events, some prefer to decompose the federation into its functional components and to identify validation referents for individual components. Then each component may be validated individually. This permits sensitivity analyses and comparisons against theory and historical data to be performed more easily. At the function-level, results may be available from other models that have comparable or greater precision and that have a prior validation history. Also at the function-level, typically more options are available for selection as possible validation referents. Once the individual functions have been validated, their interactions among M&S elements are assessed. This is accomplished by examination of the interfaces between each of the components. The V&V Agent must follow the logical threads across the interfaces to ensure that the data being received is, in fact, what was expected. While many algorithms use the same terms and produce results with the same units, they may have

underlying assumptions or approximations that unacceptably skew the results when they are utilized in subsequent processing.

The V&V Agent should seek to identify as many potential validation referents as possible for the M&S. Once the list is assembled, each potential validation referent should be analyzed to ensure it is adequate in regard to context, scope, uncertainties, etc. It helps to have multiple validation referents in order to ensure that the set of potential validation referents addresses all relevant capabilities and test objectives and incorporates the lowest uncertainty.

Insights gained from experiences with Live-Virtual-Constructive exercises since the turn of the century have helped to refine methods for assessing their validity.

Validation Referents for Humans, Software, and Systems in Distributed M&S

Validation referents for humans, software, and systems in distributed M&S are discussed below. Some aspects of human behavior representation were discussed earlier.

1. Human Factors Probably the greatest uncertainty in M&S performance is encountered in the representation of human performance. The following should be noted:

- The "average person" is not the sum of the "average" body parts. There are significant variations between the genders; in addition to size, one finds differences in reach and mobility. These have great significance when modeling the physical properties of the human body; and when simulating its motions.
- Differences in decision and reaction times have been observed between members of the Army, Navy and Air Force performing the same task in operational situations in which all had completed the same training for the task and had comparable experience. This has been ascribed to differences in the underlying command structures and operational doctrines of the services. This is a specific example of context that should be described in the referent.
- Age, experience, outside distractions and fatigue can cause significant variations in human performance. These factors must be taken into account in the simulation of Command and Control systems and in the modeling of equipment operation.

When selecting a validation referent for human performance, the first criterion is that the data sets should be as representative of the simulated population as possible. The second selection criterion is that the task represented by the data set should be representative of the action to be taken—motion or decision. Note that there can be significant differences in timing between training and exercise conditions and actual operating conditions; however, as proficiency increases, the two values will converge.

In situations in which data sets are not available from tests, textbooks and industry consensus standards can be consulted for "typical" or "standard" values for similar tasks. These data may be used to calculate a referent value for the simulation. Care should be taken to ensure that the sources selected for the referent are independent from those selected by the M&S developer for use in developing the M&S.

2. Software Software can be used as a validation referent. It may take two forms. First, a software package that has previously been accredited as representative of the system being simulated may be used as a validation referent or comparison standard for the M&S

being validated. Second, a specialty or general purpose software package may be used to construct a reference model of the system being simulated. In each case, accuracy and uncertainties associated with the software should be noted (as a standard part of referent description).

The use of accredited software packages for validation assessment is particularly applicable when complex systems are being simulated and detailed models have been developed and tested. An example of such a situation is the incorporation of a representation of the Navy's AEGIS Combat System or the Army's PATRIOT Missile System into a federation. In such a case, an M&S provided by the program office could be used as the referent for validation assessment of representation of the system used in the federation.

For notional or developmental systems, reference models can be developed using commercial software packages. Examples of such packages are COMNET® for communications systems and Extend® for process modeling. Care should be taken when constructing reference models for use as validation referents. First, the precision and accuracy of the models must be sufficient to be a satisfactory predictor of performance adequacy. Second, the model can only reflect the performances specified in the requirements and the characteristics identified to the point that the model was constructed. In this situation, the model, used as a validation referent, really represents a second solution to the system being simulated. Accordingly, many uncertainties exist. The V&V Agent should attempt to identify them and, where possible, quantify them, for their inclusion in description of the referent.

3. Real Systems Often, an operational system is represented in a federation. When this occurs, data about the system from representative tests may be used in validation assessment of the system's representation. Care must be taken to ensure that the M&S set-up representing the operational system is the same as that of the operational system at the time the data were taken. Test data about the operational system itself generally is the most authoritative source of material for the referent.

M&S Which Make Extensive Use of Aggregation

Because there is seldom appropriate test data for items represented by extreme aggregation in an M&S, the standard validation referent for M&S aggregation is based upon more detailed representations of the entities, processes, and phenomena represented in the aggregated M&S. These more detailed representations typically come from "higher fidelity" M&S. "Higher fidelity" in this context means that the actors, systems, entities, interactions, processes, or environments in the M&S are represented in more detail and at a finer granularity of resolution (for example, with individuals instead of groups as the smallest level of distinction). As noted by Bigelow and Davis, such "higher fidelity" M&S do not always give better results in terms of aggregate level behaviors.[14] Specification of the referent context is particularly important since there are often differences between the context (assumptions, conditions, etc.) of the aggregated M&S and the context of the higher fidelity M&S that can impact the appropriateness of the referent for the intended application. Use of results from "higher fidelity" M&S as referent information for aggregated M&S validation assessment brings the potential problems of Simpson's Paradox into play.[13] A validation referent should indicate how this potential problem

should be addressed if the validation referent information is structured to support piecemeal validation.

M&S with System/Hardware/Software/People in the Loop

Some validation referent considerations for this kind of M&S were addressed earlier in the subsection on Distributed M&S.

M&S can be used for understanding, describing or predicting the behavior of a single item or for items from a population of items. A flight simulator, for example, can be used to predict the likely performance of a particular pilot (or crew) in flying a particular aircraft. This is an example of a "single" item in the first sentence of this paragraph. If the same flight simulator is used to estimate performance of pilots (or crews) for a particular kind of aircraft, then it is an example of "items from a population" in the first sentence. In that case, the personnel involved may not be a good validation referent. The personnel might be aces or duds, or otherwise not representative of behavior and performance variations of the population of interest. Thorough understanding of the population of interest, its characteristics and variations, is essential in determining what should be the validation referent for this kind of M&S. Typically that issue is ignored, and often a single item is blithely assumed to be representative of the population of the item. Consideration of characteristics and variations of the item is essential in determining what should be the validation referent for this kind of M&S.

Validation referent specification for this kind of M&S should indicate whether intended use is for a single item or for items representative of a population. If it is a single item, then the relationship of that item to the validation referent should be clear. When the validation referent is the item itself, then variations in the context for the item in the M&S should be clearly delineated and their impact estimated (such as the "white knuckle" reality of real flight versus experience in a simulator, or the lack of vibration and acceleration forces in a missile seeker in a hardware-in-the-loop M&S vice one in actual missile flight). If items representative of a population are intended, then the validation referent should specify the characteristics and variations of the population and the relation of the items used in the referent to that population.

Validation Reference Dependence on M&S Application Category

Information needed for the validation referent can vary with the kind of M&S application, as illustrated by discussion of validation referents for five varieties of M&S applications. However, guidance about validation referent identification, selection, and specification/description presented earlier applies in all cases.

Computational Science and Engineering Applications

The computational science and engineering M&S community is focused on experimental data (from particular tests) or in some situations from standard benchmark cases as referents for validation assessments. In that community, "benchmarks" are usually part of verification activities. "Benchmark solutions refer either to analytical solutions, i.e., exact solutions to the partial differential equations with the specified initial conditions and boundary conditions, or to highly accurate numerical solutions. However, we believe that in the solution of nonlinear partial differential equations or solutions with discontinuities or singularities the most reliable benchmark solutions are analytical solutions. In validation activities, accuracy is measured in relation to experimental data, i.e., our best indication of reality."[15] Some in this community

reject use of the term validation for comparisons with results from other simulations (no matter how accurate those M&S results are perceived to be) or for comparison with theoretical predictions. They prefer to restrict the term validation to acceptable comparison with experimental data.

The computational science and engineering community has produced the only college/graduate-level textbook on V&V.[7] General V&V materials and that for other communities typically are a chapter in a textbook, or just conference papers and journal articles. Two professional societies related to this community have produced V&V guides for M&S.[5, 6] Both guides stress use of data from special tests (validation experiments) as the most appropriate source for the validation referent.

Engineering Level Applications

Engineering level M&S tend to have resolution at the component or sub-component level and use algorithms that describe item characteristics and behavior at that level or higher. Typically the information used as the referent in validation assessments are drawn from specifications and requirements (for future systems or items), from tests and other data (for existing systems and items), or from higher fidelity M&S (such as computational science and engineering M&S). Key aspects of validation referent specification are: identification of information sources, description of how information from different sources will be combined, quantification of variations and uncertainties in the information, and delineation of the information context.

Gaming/Training Applications

The validation referent for an M&S used in a gaming or training application depends upon the function of the application, i.e., its intended use. If the function is education, the validation referent probably can be very different from the validation referent when the function to train tank drivers in the physical tasks of driving a tank. For example, in an M&S supporting an educational endeavor, physical appearance of an object may be distorted cartoon-like to make lesson points clearer. Whereas in the M&S supporting tank driver training, realistic appearance and behavior of things is important. Generally information sources suitable as possible validation referents for description of characteristics and behavior of physical systems represented in the M&S are the same ones which might be used as referents for other kinds of M&S applications except when M&S purposes dictate unrealistic appearance or behavior. M&S used in games and training may mix human players/participants with computerized human behavior representation in the simulated environment, with human behavior representation generated by varieties of adaptive programming. Comments earlier about validation referents for M&S with adaptive programming, human behavior representation, and people in the loop apply here.

Concern about validation of computer/computer-supported games has a long history. Hermann's conclusions about game validity published in 1967 still hold today: 1) validation is always a matter of degree, which means a M&S may be more valid by some criteria than by other criteria, 2) validation cannot be separated from M&S purpose, 3) given multiple validity strategies, several of the broadly applicable criteria may be applied in a particular phase of the M&S lifecycle, and 4) human participation in the M&S significantly alters the required validation procedures. Hermann also noted that at that time only face-validation had been explored much.[16] Consequently, the M&S they were using were of unknown validity. As in

other areas, the information that might be appropriate as a validation referent depends upon the situation, and often will involve use of SMEs.

It still appears that validation referent guidance presented earlier applies in this application domain as well as elsewhere: identification of possible validation referents as a function of the situation with selection of validation referents early in the process of M&S development/assessment, specifying the validation referent in the accreditation plan (and perhaps elsewhere also) such that the validation referent is clearly and unambiguously designated using the five-fold aspect description presented earlier.

Military Theater-Level/Campaign M&S

Validation assessment of theater-level and campaign M&S is one of the most vexing problems facing V&V personnel within DoD. This is due to two characteristics. The first is the inherent complexity of the system being modeled. The second is the number of uncertainties that must be dealt with. This poses many problems in identification and selection of validation referents. Four basic methods are to address these problems. Usually these methods are used in a combination of methods.

One method is to compare results of the theater-level M&S with more detailed engagement-level M&S that address part of what the theater-level M&S addresses. Great care must be taken to ensure that the M&S employ the same context. Often the theater-level simulation and the engagement-level simulations are focused on different issues which can preclude valid comparisons of their results. When such differences do not preclude appropriate comparisons, results from the theater-level M&S may be shown to be consistent with detailed M&S that in some cases can be traced back to more elementary engineering considerations.

A second method is to disaggregate the theater-level MM&S into operational or functional areas—logistics, weapon effects, maneuver and so forth. Then the results of detailed M&S of each area can be compared with the performances found in the theater-level M&S. The validation referent (results from more detailed M&S) in this method is the same as in the previous method.

The third method examines theater-level M&S performance in terms of the comparability of its results to other M&S of greater or lesser detail and to live test and operational data.

The fourth method is reliance upon SME judgment that M&S performance complies with SME expectation.

Validation issues associated with comparing or combining information from M&S with different levels of resolution have been studied extensively.[14] It is important to specify how the possible sources of information that will be used for validation assessment of the simulation are going to be combined and used.

Although M&S in this application domain present serious V&V challenges, validation referents for such can be selected and specified in accordance with guidance presented earlier. The thorough description of the validation referent prescribed by the guidance in this Special Topic will help to ensure that the uncertainties and context of the validation referent are identified and their potential impact on the validation assessment clear.

Non-physical Applications

Non-physical applications, such as M&S related to economic theory, are different from warfighting M&S applications widely used in DoD. The primary sources of validation referent information for non-physical M&S applications are 1) theory related to the topic, or 2) historical experience or evidence related to the reality represented in the M&S. Regression techniques are frequently employed to show how M&S results correlate with the historical information. It is important that such historical information be decomposed in all areas of interest for the M&S application, and that correlations be performed in all areas to ensure that the M&S can support its intended application.

Validation referents for M&S in this application domain can be identified and specified in accordance with the guidance provided earlier.

General Applicability of Validation Referent Guidance

Five varieties of M&S and five M&S application areas were considered to see if guidance about validation referent identification, selection, and specification presented earlier were inappropriate for some of them, or if additional guidance were required for some of them. The conclusion from examining the various kinds of M&S and the various M&S applications is that the guidance presented about validation referent identification, selection, and specification applies in all cases and that no additional guidance seems to be necessary.

The quality (availability, accuracy, uncertainties, preciseness, circumstances, context, etc.) of the information selected for the validation referent will vary with M&S variety, application, and other factors, but the guidance presented about when the validation referent should be specified, where it should be documented, and the way it should be described explicitly and unambiguously (with the 5-fold aspects of its description) applies in all cases.

Validation Referent Approval

Because of the important role that validation referents play in accreditation assessment, the validation referent should be approved by the Accreditation Authority. This may be done as part of approval for the accreditation plan (which should specify the validation referent) or as a standalone item for approval by the Accreditation Authority.

Validation Referent Check-List

This section provides a check-list that can be used to ensure that a validation referent is properly identified and specified. It addresses the suggestions and guidance presented earlier in this Special Topic. The check-list is designed for those who identify and specify the validation referent.

1. Has guidance been given relative to validation referent selection and specification?

   a) From whom? M&S User, Accreditation Authority, Accreditation Agent, V&V Agent, some other authority?

   b) What does the guidance address?

   - Validation referent selection
   - Validation referent description

- Validation referent use

c) Has guidance been followed? If not, why?

2. Are needed M&S information items available in acceptable form (intended use statement, requirements, acceptability criteria, accreditation plan, V&V plan)? If not, how is needed M&S information to be developed to support validation referent identification, selection, and specification?
3. Is an adequate validation referent available? If not, has an appropriate decision been made about what should be done in the absence of an adequate validation referent?
4. Are multiple adequate validation referents available?
   a) What basis is used for selecting among them? Direction, convenience, economics, scope (portion of intended use domain covered by the possible validation referent information, best information available, broadest community acceptance, other factors), etc.?
   b) If more than one information source is used in the validation referent, how will information coherence be achieved?
5. Has the validation referent been specified appropriately with descriptions of context, domain coverage, attribute distinctiveness, quantification of parameter uncertainty, and information coherence?
6. Has validation referent dependence upon M&S variety and kind of application been addressed explicitly?
7. Has expected validation referent use in validation and accreditation assessments been defined adequately?
8. Has the validation referent been developed in accordance with specific guidance for referents for that kind of M&S?
9. Are there any validation referent issues which need to be brought to the attention of the M&S User, Accreditation Authority, Accreditation Agent, V&V Agent, or some other authority? How will the issue be brought to the attention of those who need to be aware of the issues?
10. Has the validation referent been approved by an appropriate authority? M&S User, Accreditation Authority, Accreditation Agent, V&V Agent, another?
11. Is the referent documented appropriately? In Accreditation Plan/Report and V&V Plan/Report ala MIL-STD-3022? Also in conceptual model? Elsewhere?

Validation Referent Use in M&S Validation and Accreditation Assessment

This section is different from the preceding ones. It is concerned with validation referent use instead of validation referent identification and description. It is included here because it emphasizes the importance of proper identification and description of the validation referent. Three basic situations are discussed. In the first, validation referents are data-rich (the information about the simuland for the validation referent is both abundant and accepted as pertinent and reliable). In the second, validation referents are data-poor (there is little information about simuland for the validation referent, but what data there exists is recognized as

pertinent and reliable). And in the third, the validation referent is poorly defined, poorly identified, or otherwise inadequate – the data that exist are not recognized as pertinent and reliable by all concerned.

Frequently portions of the M&S use domain will have different data quality in the validation referent for portions of the M&S intended use domain. For part of the intended use domain, there may be data-rich information suitable to serve as part of the validation referent, but there will be other parts of the M&S use domain which are data poor or for which no reliable data exist. In such situations, the validation referent will have multiple parts, with appropriate characterization of the uncertainties of information in each part in the validation referent description and specification.

This section also discusses the important issue of the relationship of the validation referent used for validation and accreditation assessment to information used for M&S development and to information used as inputs in M&S runs.

Data-rich Validation Referents

A data-rich referent is always the most desirable referent situation; especially if uncertainties about parameters of the referent are well specified throughout the entire domain of M&S intended use. Such referents provide the basis for the most reliable and most credible fidelity and validation assessments of M&S results since they provide an objective and factual basis for statistical comparison of results. However, even in situations with data-rich referents, care must be taken to ensure that the information used for the referent is truly pertinent for M&S intended use. For example, if human size is part of the referent (as might be pertinent for M&S concerned with passenger movement in a new vehicle design), one must be sure that the information about human size is recent since (at least in America) humans are bigger now than they were a couple of generations ago.

In a data-rich environment, it should be possible to separate the information used for the referent from information used to design and develop the M&S so that M&S validity and capability for reliable predictions can be assessed more robustly. The need for statistical independence between information used for M&S development and information used for validation assessment was mentioned earlier in the report and is discussed below.

Data-poor Validation Referents

In a data-poor situation, the referent is mainly theoretical by necessity. The theory may be well articulated and explicitly formulated, as in an astrophysical simulation of processes insides stars for which we have only inferred data, no *in situ* observations.[17] On the other hand, the theory may be unarticulated, as is often the case when SMEs are used as part of the referent. The SME may not make explicit how a judgment was reached or upon what evidence and logic the assessment is based. The credibility of fidelity and validation assessments based upon SME judgments in data-poor referents is always severely limited, but a thoughtful probing of SMEs (and good documentation of their insights) makes the best of this situation.

In data-poor environments, it is seldom possible to separate information used for M&S development from information used for validation assessment because all information has to be used in M&S development. Some, especially M&S personnel within the computational science and engineering community, would claim this limits M&S assessment to calibration and prohibits validation of M&S predictive capabilities. It is recommended that the validation

referent description note that validation referent information was also used for M&S development so that appropriate caveats may be associated with the validation assessments. This Special Topic does not attempt to prescribe how such caveats are determined, only that the lack of statistical independence between information used for M&S development and assessment be noted in the validation referent specification so that caveats can be indicated with the assessment as appropriate.

Inadequate Validation Referents

In the third case, the validation referent is not capable of clear identification or has recognized inadequacies (such as contradictory information, so that if M&S results agree with some of the "referent" data, they will not agree with other parts of the "referent" data). If there is no clear way to sort the validation referent information so that it is coherent and non-contradictory, then it is necessary to declare that M&S fidelity and validity assessments are not possible because of validation referent inadequacies. However, as Hodges and Dewar noted long ago, even a model which cannot be validated can have utility and value.[17] In the data-poor situation discussed above, the use of theory or SMEs as the validation referent allows one to make fidelity and validation assessments, even though the credibility of such may be low; but when the validation referent is contradictory, one cannot even make a low credibility assessment. Inadequate validation referents challenge the courage and professionalism of V&V personnel since inadequate validation referents force V&V personnel (if competent and honest) to declare to M&S sponsors, users, Accreditation Authority, and others that validation assessment is not possible because of validation referent inadequacies. Most V&V personnel are acutely aware of how often the messenger bearing bad news has been shot; many have scars that demonstrate the hazard of bearing bad news.

Validation Referent Relation to M&S Development and Run Information

Information about the reality represented in the M&S is used to develop the M&S, in running the M&S, and as the referent in validation and accreditation assessments. It is reasonable to inquire about the relationships among these three sets of information since they all deal with the same reality.

In a data-rich environment, it is most desirable that the set of information used for M&S development, the set of information used as inputs for running the M&S, and the set of information used as validation referent be statistically independent. This is particularly important when M&S results are used to predict how things will be in regions for which data are sparse or absent. The rationale for this is simple. This approach provides the highest likelihood that M&S results will be most representative of the reality represented in the M&S.

The medical community has found that conclusions from clinical trials can be vary significantly from what is believed to be more correct if appropriate care is not taken in control of such statistical issues.[8] Similar concern is needed in all M&S assessments. Many do not appreciate the issues associated with using the same information for M&S validation assessment as used for M&S development.

Concern about this general problem, and its specific related problem of not letting a modeler know experimental outcomes before the model describing the same situation is run, is abundant in the computational science and engineering community. The need for the modeler to know the

exact conditions of the experiment precisely before running the model is understood. Numerous guidelines are presented for the ways that "empirical adjustable parameters" [Roache, 1998], *knobs, dials, fudge factors* in more colloquial terminology, must be treated for M&S results to be acceptable in peer-reviewed circumstances.

In data-poor environments, statistical independence among the three sets of data may be impossible. There simply may not be enough data to permit separate data for the validation assessment referent, given what is needed to develop and run the M&S. The three data sets (development, run, and assessment referent) may even have to be identical. This condition limits what can be claimed about the validity of simulation results. One can describe with quantitative precision how well the M&S reproduces its input data, but one cannot make very meaningful comments about M&S predictive capabilities. In that regard, one may be unable to make a stronger assessment than one can with an inadequate validation referent. Candor about limitations of validation assessments in such circumstances is an important aspect of V&V professionalism. The words of statistician George Box, who at the time was the Past President of the American Statistical Association and President of the Institute of Mathematical Statistics, should not be forgotten in this respect: "all models are wrong, but some are useful."[19]

References

1. Department of Defense. *DoD Modeling and Simulation (M&S) Glossary* (Draft), March 2010.

2. Gross, David C., "Report of the Fidelity Implementation Study Group," *Proceedings of the Spring 1999 Simulation Interoperability Workshop (SIW)*, paper 99S-SIW-167A, 1999.

3. Department of Defense. DoD Instruction 5000.61: DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A). By Department of Defense. 2009.

4. Pace, Dale K., "The Referent in Simulation Development and Assessment," *Proceedings of the European Simulation Interoperability Workshop*, June 28-July 1, 2004, Edinburgh (Scotland), paper E04-SIW-024, available at http://www.sisostds.org.

5. American Institute of Aeronautics and Astronautics (AIAA), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA G-077-1998.

6. American Society of Mechanical Engineers (ASME), *Guide for Verification and Validation in Computational Solid Mechanics*, ASME V&V 10-2006, December 29, 2006.

7. Roache, Patrick J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers (Albuquerque, NM), 1998. ISBN 0-913478-08-3.

8. Kunz, Regina and Andrew D. Oxman, "The Unpredictability Paradox: Review of Empirical Comparisons of Randomized and Non-randomized Clinical Trials," *British Medical Journal*, October 31, 1998, pp. 1185-1190. Available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9794851&dopt=Abstract (accessed December 2010).

9. Pace, Dale K. and Jack Sheehan, "Subject Matter Expert (SME)/Peer Use in M&S V&V," Paper A4 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

10. Department of Defense (DoD) Standard Practice: *Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations*, MIL-STD-3022, 28 January 2008.

11. North Atlantic Treaty Organization (NATO) RTO Technical Report TR-HFN-128, *Human Behavior Representation in Constructive Simulation*, September 2009.

12. Institute of Electrical and Electronics Engineers (IEEE) Standard 1516.4™ (IEEE Std 1516.4™-2007), *IEEE Recommended Practice for Verification, Validation, and Accreditation of a Federation – An Overlay to the High Level Architecture (HLA) Federation Development and Execution Process*, 20 December 2007.

13. Pearl, Judea, *Simpson's Paradox: An Anatomy*, UCLA Cognitive Systems Laboratory, Technical Report (R-264), November 1999. Available at http://singapore.cs.ucla.edu/R264.pdf (accessed December 2010).

14. Bigelow, James H. and Paul K. Davis, *Implications for Model Validation of Multi-Resolution Modeling (MRM) and Exploratory Analysis*, RAND Corporation Report DRR-3045-AF, April 2003.

15. Oberkampf, William L., Timothy G. Trucano, Charles Hersch, "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," Paper B1 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

16. Hermann, Charles F., "Validation Problems in Games and Simulations with Special Reference to Models of International Politics," *Behavioral Science*, Vol. 12, pp.216-231, 1967.

17. Calder, Alan, Jonathan Dursi, Bruce Fryxell, and 8 others, "Validating Astrophysical Simulation Codes," *Computer Science Engineering*, Vol. 6, Issue 5, September 2004, pp. 10-20.

18. Hodges, J. S. and J. A. Dewar, *Is It You or Your Model Talking? A Framework for Model Validation*, RAND Report, R-4114-AF/A/OSD, 1992.

19. Box, G. E. P., "Robustness is the Strategy of Scientific Model Building" in *Robustness in Statistics*, eds., R.L. Launer and G.N. Wilkinson, Academic Press, 1979, p. 202.