Towards A Probabilistic Interpretation of Validity for Simulation Models

James J. Nutaro
Oak Ridge National Laboratory
One Bethel Valley Road
Oak Ridge, TN, 37831
nutarojj@ornl.gov

Bernard P. Zeigler
RTSync Corp. and Arizona
Center for Integrative Modeling
and Simulation
Phoenix, AZ
zeigler@rtsync.com

ABSTRACT
We propose an approach to addressing questions about the validity of a simulation model and we develop this approach in detail with respect to two specific objectives. The first objective is to decide when a successful test conducted on a model of a system justifies increased confidence in the system. The second objective is to decide when a model may be replaced by another model such that successful tests on the replacement increase our confidence in the system. The basis of our approach is a probabilistic interpretation of validation and we give specific definitions in terms of the response function of the system and its models. Within this approach, we derive conditions under which confidence in a model justifies it as a surrogate in testing and in similar conditions under which a model may be replaced by another.

Author Keywords
Validation; verification; modeling and simulation; experimental frames; intended use

ACM Classification Keywords
I.6.4 MODEL VALIDATION AND ANALYSIS

INTRODUCTION
Extending the work of Zeigler [1,2] in developing a theory and framework for modelling and simulation, Traore and Muzy [3] formalized experimental frames to describe the experimental contexts in which a simulation model is a valid representation of some real or imagined system. Implicit in this formalization is the idea that we can state precisely the experimental frames in which a model is valid. Though in principle this description exists, it is more common to assert an intended use for the model and take this as the basis for deciding what experiments will yield valid results. The key difference between experimental frames and intended use is that the experimental frame is crisply defined: an experimental plan either falls within its scope or does not. Intended use is an imprecise statement about an incompletely understood set of experimental frames that are meant, but have not been conclusively proven, to be valid scopes for the use of the model.

This distinction between experimental frames and intended use is immediately relevant to two related problems. The first problem is to what extent a successful test performed on a model informs us about the likely outcome of the same test performed on the real or imagined system. This problem emerges routinely in areas of engineering where simulation is used to supplant live testing and training: examples include simulated testing of nuclear weapons [4] and simulators used for training pilots and tank crews [5]. The second problem is to what extent one model can be replaced with another within a given experimental frame. The desire to replace one simulation model for another also occurs frequently in simulations of engineered systems. A common motivation for replacement is that one model is more computationally tractable than another and thereby enables more experiments within a fixed budget of time and money. Another motivation is software obsolescence, which happens when the supplier of a simulation model ceases to support an old version of its software and, in effect, forces a transition to some newer version.

Regardless of the motivation, a primary justification for using a model or its proposed replacement is the model's intended use. A central question is whether the intended uses of the model and its valid uses coincide in the experimental frames of interest. The direct answer to this question comes from extensive validation. If the model is of a real system, then this involves very costly and time consuming comparisons of the model's behavior with that of the real system. If the model is of an imagined system, a situation that occurs early in an engineering process before prototypes have been built, then validation by direct comparison is impossible and we must rely on weaker evidence (see, e.g., the discussion by Sargent [6]).

We propose that questions of a model's suitability as a surrogate for testing and questions about replacing one model with another can be cast as probabilistic inquiries. One outcome of this approach is that we can determine

conditions for which a successful test on a model increases our confidence that the same test on the real or imagined system will also be successful. Another outcome is theoretical support for the intuitively appealing idea of validating one simulation model by comparing it with another that is known to be valid. In this case, we provide conditions for which a successful test with the replacement model increases our confidence that the same test on the real or imagined system will also be successful.

The basis for our approach is the response functions of the models and system. The response function is sufficient to capture the initial conditions of a test, the applied stimulus, and the observed behavior. Consequently, it describes the essential features of a model and system when asking about validity with respect to an experimental frame; that is, whether the model is essentially indistinguishable from the system (real or imagined) with regards to the input, output, and initial states that constitute a set of tests or experiments. Within this very simple setting, we interpret validity as a probabilistic statement about the intersection of response sets within an experimental frame. These probabilistic statements are naturally tied to the intended use of a model when we interpret probabilities as measures of confidence.

Our approach could be usefully cast in any number of frameworks for statistical inference, though we do not attempt to do so here. Nonetheless, a clear link to these can be seen in our underlying assumption of fixed space of behaviors for the model, system, and experimental frame and that limited samples of this space are used to infer the likelihood (or, rather, the increase in likelihood) that some future sample will fall within an intersection of interest. By putting our ideas forward in a rudimentary fashion, we hope to make the underlying assumptions transparent and reserve for future work applying more powerful methods of statistical inference.

VALIDATION AND INTENDED USE

We describe a system with a set of initial states Q, a set X of input trajectories, and a set Y of output trajectories. The system's response function $\rho(q, x) = y$ maps each initial state and input trajectory to an output trajectory. Consider the special case of a large system $< Q, X, Y, \rho >$ and a small system $< Q', X', Y', \rho' >$ such that $Q' \subseteq Q$, $X' \subseteq X$, and $Y' \subseteq Y$. The small system is embedded in the large system if

$$\rho(q', x') = \rho'(q', x')$$

for all $q' \in Q'$ and $x' \in X'$. (This is a special case of an I/O functional homomorphism from the large system onto the small system; see [2,7]).

In practice, we are concerned with experiments or tests that encompass just some of the possible or desirable behaviors of the large system. These tests of interest are the experimental frame, which for our purposes is described by a subset E of $Q \times X \times Y$. The functions ρ and ρ' define sets of triples of input, output, and initial state that are realizable by the large and small system respectively, and we use ρ' and ρ to indicate both the function and the corresponding set as the intended interpretation will be clear from the context. If tests in E are to be feasible for the large system then we would require $E \subseteq \rho$, and for the small system to be useful as a model in these tests we would require $E \subseteq \rho'$. This definition of the experimental frame can be interpreted as the result of applying an experimental frame as described by Traore and Muzy [3] to the system; the distinction is not material to what follows and so, with some abuse of the term, we use experimental frame to refer to the set E.

If the experimental frame is a subset of ρ' and the above embedding exists then any behavior we observe in the model will also be observed in the system, but the reverse need not be true. Consequently, if we conduct a test on the model by placing it into an initial state, applying an input trajectory, and observing the response, then we know that the same test conducted on the system will produce the same result. This assurance, or belief in this assurance, is the basis for using models in tests and for design. With complete knowledge of the experimental frame, system, and model, the question of validation is reduced to showing that the model is embedded in the system for all objects in the experimental frame. If it is then tests of the model anticipate the outcome of tests of the system. Moreover, we could answer the question of a model M' being a suitable replacement for a model M: if M is embedded in the system and M' is embedded in M, then M' is also embedded in the system and we would conclude that M' is a valid replacement for M.

In practice this is impossible because we do not have a precise, accurate, or complete definition of the experimental frame, system, and model. We do not know with certainty what behaviors can be realized by the models or the system, and we are incapable of asserting how either will respond to all relevant pairings of input and initial state. Indeed, if we knew the answers to these questions then we would have no need for the models as engineering tools. Worse still, we cannot completely characterize the experimental frame. Stimulus to the model is very likely to come from another poorly characterized system or model, and it might even be a function of output from the model itself, such as when two models are connected in a feedback arrangement.

Intended use

Faced with these uncertainties, we rely on what a model is intended to do rather than what a model is known to do. The more we know about a model and system, the greater our confidence in the statement of intent, but the intended valid uses of the model are almost always a superset of the known valid uses. From this perspective, we can interpret intended use as the belief that the model is valid for a set of experiments. A concrete definition of this belief is a conditional probability that the system will exhibit some particular behavior given that the model exhibits this behavior. To develop this definition, we proceed as follows.

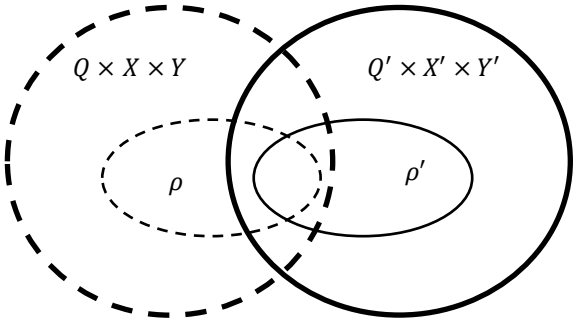

Figure 1. Unions and intersections of the set of possible behaviours of the system (thick dashed line), actual behaviours of the system (thin dashed line), possible behaviours of the model (thick solid line), and actual behaviours of the model (thin solid line).

Let $S =< Q, X, Y, \rho >$ be the real (or imagined) system and $M =< Q', X', Y', \rho' >$ be another system (e.g., a computer program) that we wish to use as a model of S. By calling M a model of S we express a hope that $Q' \subseteq Q$, $X' \subseteq X$, $Y' \subseteq Y$, and $\rho' \subseteq \rho$ but we do not know that this is the case. Moreover, we do not necessarily know what elements constitute ρ' and ρ; again, if we did then we would have no need for models and tests.

It is possible that there are elements in $Q \times X \times Y$, where the system exists, that are not in the set $Q' \times X' \times Y'$, where the model exists, and vice versa. Therefore, to examine where the model and system intersect it is necessary to consider the set $(Q \times X \times Y) \cup (Q' \times X' \times Y')$ of all possible behaviours. This is illustrated in Figure 1. Suppose we select a test at random from this set. As a technical convenience, we will consider the case where all sets under consideration are finite. The probability $P(M)$ that M will exhibit the selected behavior and pass the test is the probability that the test exists in the set ρ', and the probability $P(S)$ that S will pass the test is defined in the same way. These probabilities are

$$P(S) = \frac{|\rho|}{|(Q \times X \times Y) \cup (Q' \times X' \times Y')|} \quad \text{and}$$

$$P(M) = \frac{|\rho'|}{|(Q \times X \times Y) \cup (Q' \times X' \times Y')|} \quad .$$

The probability $P(MS)$ that both M and S will pass the test is

$$P(MS) = \frac{|\rho' \cap \rho|}{|(Q \times X \times Y) \cup (Q' \times X' \times Y')|}$$

and the conditional probability $P(S|M)$ that S passes given M passes is

$$P(S|M) = \frac{P(MS)}{P(M)} = \frac{|\rho' \cap \rho|}{|\rho'|} .$$

The probability $P(S|M)$ is the fraction of the total set of possible tests for which there is an embedding of M in S (admitting the possibility that this space may be empty), and this is naturally interpreted as the likelihood that a successful test conducted on the model is representative of a successful test conducted on the system. Indeed, if M is contained in S then $\rho' \subseteq \rho$ and

$$P(S|M) = \frac{|\rho' \cap \rho|}{|\rho'|} = \frac{|\rho'|}{|\rho'|} = 1 .$$

In practice we cannot determine $P(S|M)$ with certainty. Hence, our estimate of this probability is a statement about our confidence in the validity of the model with respect to the system in a given experimental frame. Stated another way, our estimate of $P(S|M)$ is a quantification of *our confidence in the model's intended use.*

Conditions for valid use

Using this probabilistic interpretation of validity, we cast our two primary questions as ones of revising conditional probabilities based on the outcome of an experiment. To recall, the first of these questions concerns using tests on a model as a substitute for tests on the system: if we perform a successful test on the model, under what conditions does this increase our confidence that the same test on the system will be successful? The second question concerns replacement: if we replace one model with another, under what conditions do successful tests with the replacement increase our confidence of a successful test with the system?

As a prelude to the first question, we may consider the experimental frame $E \subseteq Q \times X \times Y$ to be a specification of desired behaviors (e.g., a set of requirements) and the system S to be a model of the desired behavior. For example, S could be a design concept or prototype built to satisfy a set of requirements. We then ask what is the probability that S exhibits some particular behavior given that behavior is in E. The design or prototype is valid (e.g., it satisfies its requirements) if $P(S|E) = 1$, which means that $E \subseteq S$ any test of the system for a behavior in E will be successful. If $P(S|E) < 1$, then there are some tests that the design or prototype will fail; that is, there are behaviors in E that are not realized by S.

A simulation model M of S may be used for a test that is impractical to conduct on S. The conditional probability $P(M|E)$ is the probability that M will pass the test and $P(S|M)$ is the probability that S will pass given that its model passed. Prior to the test on M, we have an estimate $P(S|E)$ of how likely the system is to pass the test. Then we conduct the test on M and obtain a positive result. Should our estimate of $P(S|E)$ be increased?

To describe conditions under which this is true, we first define a positive outcome of a test in terms of how it revises our estimates of probability. A successful test that a behavior in E is also in M increases our estimate of $P(M|E)$ by some $k > 0$ and, because $P(\bar{M}|E) = 1 - P(M|E)$, it reduces our estimate of $P(\bar{M}|E)$ by the same amount. The test does not give any information about behaviours that are not in E, and so our estimates of conditional probabilities dependent on $\bar{E}$ are unchanged by the test. Similarly, the test gives us no new information about the relationship between the model and system, and so our estimates of $P(S|M)$, $P(M|S)$, and so forth are unchanged.

An example will help to clarify this model of a test. Suppose the set of all possible behaviors (our domain of discourse) is $\{b_1, b_2, b_3, b_4\}$, $M = \{b_2, b_3\}$, $E = \{b_2, b_3, b_4\}$, and our initial estimate of $P(M|E) = 0$. That is, we assume M and E do not intersect. Now we select b_4 from E and perform a test for its inclusion in M. Of course, this test is negative and so our estimate of $P(M|E)$ is unchanged. Next we select b_3 from E and perform a test for its inclusion in M. This test is positive, indicating that $ME \supseteq \{b_3\}$ and so we revise our estimate upward to 1/3. Lastly, we select b_2 from E and perform a test for its inclusion in M. Again, the test is positive and (because this is the last element in E) we know that $ME = \{b_2, b_3\}$ and that $P(M|E) = 2/3$. For the sake of technical simplicity, we assume that such a process of upward revision always converges to the actual conditional probability.

The following probabilities are essential to what follows:

$P(\bar{S}|E) = 1 - P(S|E)$, the probability that S will fail a test, and

$P(M|\bar{S})$, the probability that M has a behavior that S does not

and so is the following lemma.

Lemma 1: If $P(M|S) > P(M|\bar{S})$ then $P(S|M) > P(S|\bar{M})$.

Proof: From Bayes' Theorem

$$P(S|M) = \frac{P(M|S)P(S)}{P(M|S)P(S) + P(M|\bar{S})P(\bar{S})} = \frac{P(S)}{P(S) + \frac{P(M|\bar{S})}{P(M|S)}P(\bar{S})}, \text{ and}$$

$$P(S|\bar{M}) = \frac{P(\bar{M}|S)P(S)}{P(\bar{M}|S)P(S) + P(\bar{M}|\bar{S})P(\bar{S})} = \frac{P(S)}{P(S) + \frac{P(\bar{M}|\bar{S})}{P(\bar{M}|S)}P(\bar{S})} = \frac{P(S)}{P(S) + \frac{1 - P(M|\bar{S})}{1 - P(M|S)}P(\bar{S})}.$$

It follows from the hypothesis that $P(M|\bar{S})/P(M|S) < 1$, $\big(1 - P(M|\bar{S})\big)/\big(1 - P(M|S)\big) > 1$, and so $P(S|M) > P(S|\bar{M})$, which completes the argument. ■

We can now show that our estimate of $P(S|E)$ is increased if $P(M|\bar{S}) < P(M|S)$; that is, if the model is more likely to behave like the system than not. This condition is satisfied when M is embedded in S because then $P(M|\bar{S}) = 0$ and $P(M|S) = 1$. Otherwise, we must resort to subject matter experts to justify or refute this assumption.

Proposition 1: If $P(M|\bar{S}) < P(M|S)$ then the posterior probability $P^*(S|E)$ that S will pass a test given its model M has passed is greater than the prior probability $P(S|E)$ that S will pass.

Proof: The effect of the successful test is to raise our estimate of $P(M|E)$ by k to $P^*(M|E)$ and to lower our estimate of $P(\bar{M}|E)$ by k to $P^*(\bar{M}|E)$. Therefore, we may write

$P^*(M|E) = P(M|E) + k$, and

$P^*(\bar{M}|E) = 1 - P^*(M|E) = 1 - [P(M|E) + k] = P(\bar{M}|E) - k$.

By definition

$P(S) = P(S|E)P(E) + P(S|\bar{E})P(\bar{E}) = P(S|M)P(M) + P(S|\bar{M})P(\bar{M})$,

$P(M) = P(M|E)P(E) + P(M|\bar{E})P(\bar{E})$, and

$P(\bar{M}) = P(\bar{M}|E)P(E) + P(\bar{M}|\bar{E})P(\bar{E})$.

By rearrangement and substitution in the above we obtain

$$P(S|E) = \frac{P(S|M)[P(M|E)P(E) + P(M|\bar{E})P(\bar{E})]}{P(E)} + \frac{P(S|\bar{M})[P(\bar{M}|E)P(E) + P(\bar{M}|\bar{E})P(\bar{E})] - P(S|\bar{E})P(\bar{E})}{P(E)}$$

and recalling our definition of the test our revised belief is

$$P^*(S|E) = \frac{P(S|M)[P^*(M|E)P(E) + P(M|\bar{E})P(\bar{E})]}{P(E)} + \frac{P(S|\bar{M})[P^*(\bar{M}|E)P(E) + P(\bar{M}|\bar{E})P(\bar{E})] - P(S|\bar{E})P(\bar{E})}{P(E)}$$

Subtracting and cancelling terms gives

$$P^*(S|E) - P(S|E) = P(S|M)[P^*(M|E) - P(M|E)] + P(S|\bar{M})[P^*(\bar{M}|E) - P(\bar{M}|E)]$$

Substituting $P^*(M|E) = P(M|E) + k$ and $P^*(\bar{M}|E) = P(\bar{M}|E) - k$ gives us

$$P^*(S|E) - P(S|E) = k[P(S|M) - P(S|\bar{M})]$$

Because $k > 0$ it is sufficient for $P(S|M) > P(S|\bar{M})$ to conclude $P^*(S|E) > P(S|E)$. That $P(S|M) > P(S|\bar{M})$ follows from the hypothesis of the Proposition and Lemma 1. ■

The above proposition is about the effect of a test given our belief that the model behaves like the system. From the proof of this proposition we immediately have the following proposition describing the effect of a test given our belief that the system behaves like the model.

Proposition 2: If $P(S|\bar{M}) < P(S|M)$ then the posterior probability $P^*(S|E)$ that S will pass a test given its model M has passed is greater than the prior probability $P(S|E)$ that S will pass.

Now consider the second problem of replacing the model M with the model M'. Referring to the above proposition, if $P(M'|\bar{S}) < P(M'|S)$ then a successful test conducted on M' increases our confidence in S. Lacking explicit knowledge of how M' relates to S, we can rely on information about the relationship between M and S and between M' and M. This leads to Proposition 3 concerning the use of M' as a substitute for M.

Proposition 3: If $P(M'|\bar{M}) < P(M'|M)$ and $P(M|\bar{S}) < P(M|S)$ then the posterior probability $P^*(S|E)$ that S will pass a test given the replacement model M' has passed is greater than the prior probability $P(S|E)$ that S will pass.

Proof: It follows from Proposition 1 that a successful test of M' raises our estimate of $P(M|E)$ and in this respect is indistinguishable from a successful test of M. Because $P(M|\bar{S}) < P(M|S)$ we also raise our estimate of $P(S|E)$. ■

In the above we have not considered any information about the likelihood of obtaining a specific result from the model and system. Lacking such information, we focused merely on how the sets of realizable behaviors and a set of desired tests might intersect. If we do have (or can guess) detailed information about the behaviors of model and system then it is possible to prove a form of Proposition 2 that considers each behavior individually.

To pursue this, we will restrict the domain of discourse to a set containing the single pair q, x of initial state and stimulus and the outputs $y_1, \ldots, y_n$ such that just the response y_1 is considered passing a test. Further suppose that there are probabilities $P_M(y)$ that $\rho_M(q,x) = y$, $P_S(y)$ that $\rho_S(q,x) = y$, and $P(y_j; y_k)$ that $\rho_S(q,x) = y_j$ given that $\rho_M(q,x) = y_k$. A successful test of model raises our estimate of $P_M(y_1)$ by k and reduces our estimate of $P_M(y_i)$, $i > 1$, by k_i such that $k_2 + \cdots + k_n = k$. Analogous to Proposition 2 we have the following.

Proposition 4: If $P(y_1; y_1) > \sum_{i \neq 1} P(y_1; y_i)$ then the posterior probability $P_S{}^*(y_1)$ that S will pass the test given the model M has passed is greater than the prior probability $P_S(y_1)$ that S will pass.

Proof: By definition

$$P_S(y_1) = P(y_1; y_1)P_M(y_1) + \sum_{y_i \neq y_1} P(y_1; y_i)P_M(y_i)$$

and

$$P_S{}^*(y_1) = P(y_1; y_1)[P_M(y_1) + k] + \sum_{i \neq 1} P(y_1; y_i)[P_M(y_i) - k_i]$$

Taking the difference yields

$$P_S{}^*(y_1) - P_S(y_1) = P(y_1; y_1)k - \sum_{i \neq 1} P(y_1; y_i)k_i \geq k\left[P(y_1; y_1) - \sum_{i \neq 1} P(y_1; y_i)\right]$$

If $P(y_1; y_1) > \sum_{i \neq 1} P(y_1; y_i)$ then this difference is positive. ■

To illustrate this proposition consider the case of just two behaviors. The condition of the proposition is $P(y_1; y_1) > P(y_1; y_2)$. This implies $P(y_1; y_1)/P(y_1; y_2) > 1$, which is true only if $P(y_1; y_1) > 1/2$.

Proposition 4 is closely related to Proposition 1 and 2. Because q, x is fixed, it will be convenient to define sets in terms of outcomes. The relevant sets appearing in Proposition 2 are $S = \{y \mid \rho_S(q,x) = y\}$ and $E = \{y_1\}$. Because E has just a single element, our estimate of the probability $P(S|E)$ is simply $P_S(y_1)$ and similarly $P^*(S|E) = P_S{}^*(y_1)$. From this and the proofs of Propositions 1 and 4 it follows that

$$k[P(S|M) - P(S|\bar{M})] = P(y_1; y_1)k - \sum_{y_i \neq y_1} P(y_1; y_i)k_i \geq k\left[P(y_1; y_1) - \sum_{i \neq 1} P(y_1; y_i)\right]$$

This proves the following pair of corollaries.

Corollary 1: If $P(y_1; y_1) > \sum_{y_i \neq y_1} P(y_1; y_i)$ then $P(S|M) > P(S|\bar{M})$.

Corollary 2: $P(S|M) > P(S|\bar{M})$ if, and only if, $P(y_1; y_1) > \left[\sum_{y_i \neq y_1} P(y_1; y_i)k_i\right]/k$.

Hence, if the condition of Proposition 3 is satisfied, then so are the conditions of Proposition 2 and, through Lemma 1,

Proposition 1. Moreover, the conditions of Propositions 2 and a slightly stronger condition for Proposition 4 are interchangeable: satisfaction of one implies satisfaction of the conditions for the other.

AN ILLUSTRATIVE EXAMPLE

The model proposed by Davis and Anderson [8] to demonstrate a problem of time management in distributed simulations provides a nice illustration of these propositions. In this model there are two gunfighters: A and B. They meet in the street, hand on gun, and wait for an arbitrator to say "draw". Then the gunfighters draw pistols and fire. For each there is a delayed reaction between the command to draw and taking the shot, another delay as the round travels to its target, and then the bullet hits or misses.

The complete set C of possible tests that we could conduct on this model is as follows. The gunfighters are always alive at the start of an encounter (this is the only initial state). We consider two possible delays in the reaction of each gunfighter, two possible travel times for each round, and each shot can either hit or miss its target. These options give 64 sequences of events that we can inject into the model. An encounter has four possible outcomes: A dead and B alive, A alive and B dead, both dead, or both alive. Hence, the total set of possible tests that we could conduct has 256 elements.

The gunfight "system" S realizes only the subset of these that satisfies its logical constraints. For example, suppose the possible travel times are 0.05 seconds and 0.06 seconds, and the possible reaction delays are 0.14 seconds and 0.2 seconds. The behaviors that S realizes are as follows:

1. 16 cases where A misses and B misses, and these always result in A and B alive;
2. 16 cases where A hits and B misses, and these always results in A alive and B dead;
3. 16 cases where A misses and B hits, and these always results in A dead and B alive;
4. 4 cases where A and B are both are killed;
5. 6 cases where A and B would have both hit, but B is faster and kills A; and
6. 6 cases where A and B would have both hit, but A is faster and kills B.

From this enumeration we conclude that $P(S) = 64/256 = 1/4$.

To illustrate Proposition 1 our first model M of S will be a discrete event simulation that assumes the gunfighters never miss. This simulation ignores hit and miss information supplied for the gunfighters, always replacing this data with hits. All 16 combinations of the total time from "draw" to impact are listed in Table 1 along with the outcome of the simulated encounter. With this data we calculate $P(M) = 16/256 = 1/16$. The model agrees with (intersects) S for these 16 cases (i.e., the last three items in the above list), hence $P(MS) = P(M)$ and $P(M|S) = 1/4$. The set $\bar{S} = C - S$ has 192 elements, and of these the following also occur in M:

1. 3×4 cases where A, B, or both miss, but A and B are both killed;
2. 6 cases where A hits and B misses, but B is faster and kills A;
3. 6 cases where A misses and B misses, but B is faster and kills A;
4. 6 cases where A misses and B hits, but A is faster and kills B;
5. 6 cases where A misses and B misses, but A is faster and kills B.

With this data we calculate $P(\bar{S}) = 192/256 = 3/4$, $P(M\bar{S}) = 36/256 = 9/64$, and $P(M|\bar{S}) = 3/16$. Proposition 1 shows that the model is a suitable surrogate for S because $P(M|S) = 0.25 > 0.1875 = P(M|\bar{S})$.

To illustrate proposition 2 suppose that the discrete event simulation is not available to us (e.g., it would be too expensive to construct) and instead we reuse existing models of the two gunfighters. These are inter-connecting such that they exchange data at the end of every 0.025 second time step. The effect of this strategy is to transform the time from "draw" to a round striking its target into just three possibilities:

1. The round arrives at step $\lceil (0.14 + 0.05)/0.025 \rceil = \lceil (0.14 + 0.06)/0.025 \rceil = 8$;
2. The round arrives at step $\lceil (0.2 + 0.05)/0.025 \rceil = 10$;
3. The round arrives at step $\lceil (0.2 + 0.06)/0.025 \rceil = 11$.

This new model M' agrees with M in 14 out of the 16 cases listed in Table 1, and so $P(MM') = 14/256 = 7/128$ and

$$P(M'|M) = \frac{P(MM')}{P(M)} = \frac{7}{128}\frac{16}{1} = \frac{7}{8}.$$

The set $\bar{M} = C - M$ has 240 elements and $P(\bar{M}) = 240/256 = 15/16$. Two elements of M' are in $\bar{M}$ so that $P(M'\bar{M}) = 1/120$ and

$$P(M'|\bar{M}) = \frac{P(M'\bar{M})}{P(\bar{M})} = \frac{1}{120}\frac{16}{15} = \frac{2}{225}.$$

With Proposition 2 we conclude that M' is an acceptable replacement for M because $P(M|S) > P(M|\bar{S})$ and $P(M'|M) = 0.875 > 0.00889 \approx P(M'|\bar{M})$.

To illustrate a negative result for both propositions we consider a model L for which the outcome is always that both gunfighters live. Like M, the model L contains 64 distinct behaviours. The relevant probabilities are $P(LS) = 1/16$, $P(L\bar{S}) = 48/256 = 3/16$, and $P(L|S) = P(L|\bar{S}) = 1/4$. From Proposition 1 we conclude that L is not a suitable surrogate for S. We can also calculate $P(LM) = 0$, $P(L\bar{M}) = 1/16$, and $P(L|M) = 0 < P(L|\bar{M}) = 1/15$.

From Proposition 2 we conclude that L is not a suitable replacement for M (of course, the prior result of $P(L|S) = P(L|\bar{S})$ was sufficient to draw this conclusion).

A exact	A time stepped	B exact	B time stepped	Exact outcome	Time stepped outcome
0.19	8	0.19	8	Both killed	Both killed
0.20	8	0.19	8	B wins	Both killed
0.25	10	0.19	8	B wins	B wins
0.26	11	0.19	8	B wins	B wins
0.19	8	0.20	8	A wins	Both killed
0.20	8	0.20	8	Both killed	Both killed
0.25	10	0.20	8	B wins	B wins
0.26	11	0.20	8	B wins	B wins
0.19	8	0.25	10	A wins	A wins
0.20	8	0.25	10	A wins	A wins
0.25	10	0.25	10	Both killed	Both killed
0.26	11	0.25	10	B wins	B wins
0.19	8	0.26	11	A wins	A wins
0.20	8	0.26	11	A wins	A wins
0.25	10	0.26	11	A wins	A wins
0.26	11	0.26	11	Both killed	Both killed

Table 1. Tests and outcomes in the discrete event and discrete time simulations of a gunfight.

In the above examples we had complete knowledge of the model and system and could calculate the relevant probabilities directly. To see how Proposition 1 applies when our information is incomplete, suppose that we only have the estimates $P(M|S) \approx 0.2 > P(M|\bar{S}) \approx 0.19$ and $P(M'|M) \approx 0.9 > P(M'|\bar{M}) \approx 0.01$ to relate the models and system and that for the set of tests $E \subset C$ in which A and B always hit we estimate $P(S|E) \approx 0.75$. We select a test in E that has not been tried before, apply that test to M, and obtain a positive result, thereby increasing our confidence in the model relative to E by $k = 1/|E| = 1/64 \approx 0.016$. From the proof of Proposition one, we see that $P(S|E)$ increases by $[P(S|M) - P(S|\bar{M})]/|E|$, which could be calculated using Lemma 1 if we had precise knowledge of the model and system (i.e., by using the probabilities determined in the prior examples). In fact, doing this shows us $[P(S|M) - P(S|\bar{M})]/|E| \approx 0.00113$. An example of Proposition 2 can be constructed in exactly the same way.

CONCLUSION

We have defined validity through a probabilistic interpretation of the modeler's confidence that a model's intended use and actual scope for valid experimentation overlap. With this approach we answer two questions.

1) Does a successful test on a model justify increased confidence that the same test on the system will be successful? The answer to this question is yes if our confidence in the model is sufficiently great, with sufficient being satisfaction of $P(M|S) > P(M|\bar{S})$.

2) If I replace one model with another, does a successful test with the replacement justify increased confidence in the system? Again, the answer to this question is yes if we are confident that model and its replacement are sufficiently similar, with sufficient being satisfaction of $P(M'|M) > P(M'|\bar{M})$.

The conditions for a positive answer to the first question are intuitively appealing because they provide a definition of "good enough", which is what most validation, accreditation, and certification activities strive for but have difficulty quantifying (see, e.g., the discussion of credibility and intended use by Balci et. al. in [9-11]). The conditions for a positive answer to the second question provide probabilistic justification for the practice of validating a new model through tests against another, trusted model (see, e.g., the discussion of validation by Sargent [6]).

The key challenge to applying the proposed propositions is obtaining reliable estimates for the necessary probabilities. One possible solution to this problem is to include these estimates in the model accreditation or certification process, which implicitly asserts that $P(M|S)$ is high and $P(M|\bar{S})$ is low. Methods for generating specific probabilities as part of accreditation or certification is a promising avenue for future research, which might build upon or be incorporated into existing processes such as those proposed by Balci [11] and Gass [12].

Answers to the above questions are important for using models in engineering, but they do not encompass all problems of practical interest. More generally, we would like a quantitative relationship, probabilistic or otherwise, between systems and models that allows us to infer i) if one system extends another, that some property of the simpler system is preserved (the problem of upward preservation, see e.g., [13,14]); ii) if the one system simplifies another, that some property of the more complex system is preserved (the problem of downward preservation, see e.g., [15,16]); and iii) that a composition of two systems is suitable for some purpose (the problem of model composition, see e.g. [8]). In practice, these problems are

dealt with today by statements about intent, confidence in the veracity of a stated intent, and a process of informal inference supported by validation and verification. In this sense, the more difficult problems (i-iii) are similar to the questions addressed here and may also be amenable to a probabilistic framing that is useful and precise.

REFERENCES

1. Zeigler, B.P. *Theory of Modeling and Simulation (1st ed.)*. Academic Press, 1976.
2. Zeigler, B.P., Praehofer, H., and Kim, T.G. *Theory of Modeling and Simulation (2nd ed.)*. Academic Press, 2000.
3. Traoré, M.K. and Muzy, A. 2006. "Capturing the dual relationship between simulation models and their context." Simulation Modeling Practice and Theory, 14, 2 (2006), 126-142.
4. Bennett, L., Jablonski, D., Lansrud-Lopez, B., Scott, J., and Webster, B. "Part 2: Second-Generation Designers". National Security Science, Feb (2014), 27-33.
5. Dewar, J.A., Bankes, S.C., Hodges, S., Lucas, T.W., Saunders-Newton, D., and Vye, P. *Credible Uses of the Distributed Interactive Simulation (DIS) System*. MR-607-A, RAND Corporation, 1996.
6. Sargent, R.G. "Verification and validation of simulation models." In Proceedings of the Winter Simulation Conference (WSC '11), S. Jain, R. Creasey, J. Himmelspach, K. P. White, and M. C. Fu (Eds.), (2001), 183-198.
7. Mesarovic, M.C. and Takahara, Y. *Abstract systems theory*. Springer-Verlag, 1989.
8. Davis, P.K. and Anderson, R.H. *Improving the Composability of Department of Defense Models and Simulations*. MG-101-OSD, Rand Corporation, 2004.
9. Balci, O. "How to Assess the Acceptability and Credibility of Simulation Results." In Proc. 1989 Winter Simulation Conf., edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, (1989), 62-71.
10. Balci, O. and Ormsby, W.F. "Well-defined intended uses: an explicit requirement for accreditation of modeling and simulation applications." Proceedings of the Winter Simulation Conference, (2000), 849-854.
11. Balci, O. "A methodology for certification of modeling and simulation applications." ACM Transactions on Modeling and Computer Simulation, 11, 4 (2001), 352-377.
12. Gass, S.I. "Model accreditation: A rationale and process for determining a numerical rating," European Journal of Operational Research, 66, (1993), 250-258.
13. Sierocki, I. "A Note On Structural Inference In Systems Theory." International Journal of General Systems, 13, 1 (1986), 17-22.
14. Saadawi, H. and Wainer, G. "Principles of Discrete Event System Specification Model Verification." SIMULATION, 89, 1 (2013), 41-67.
15. Foo, N. "Stability Preservation under Homomorphisms." IEEE Transactions on Systems, Man, and Cybernetics, 7, 10 (1977), 750-754.
16. Foo, N. "Homomorphisms in the Theory of Modeling." International Journal of General Systems, 5, 1 (1979), 13-16.