# A Risk-Based Validation Approach for Irregular Warfare Models

**Lisa Jean Bair, Eric W. Weisel**
**WernerAnderson, Inc.**
**Gloucester, VA**
**lbair@werneranderson.com, eweisel@werneranderson.com**

**Richard Brown**
**US Army TRADOC Analysis Center – Monterey**
**Monterey, CA**
**richard.f.brown1@us.army.mil**

## ABSTRACT

The defense community requires a robust capability to represent and analyze the Irregular Warfare (IW) environment across the range of tactical, operational, and strategic levels of warfare. In support of this need, TRAC Monterey is developing a prototype capability that credibly represents ground forces conducting Counter Insurgency (COIN) operations focusing on the relationships and interactions with a population of interest. While understanding the validity of the M&S of physics-based systems for a given use is well-understood and physics-based combat models have a long history of use, methods and tools for assessing the validity of M&S in an IW environment are not readily available. In recognition of this need, we have developed a measurable, repeatable method for assessing, understanding, and describing the risk of using an M&S for analysis. This approach is unique in that we have developed risk measures for using a model or simulation for a specified application and criteria for assessing the risk of using a model or simulation based on consequence, error, and validation process. To exercise validation methods developed in this effort, a validation of the Cultural Geography Model (CGM) was performed to assess the appropriateness of the CGM representation of the Conflict Ecosystem conceptual model; the Social Network Representation within the CGM and the whether this representation is generalizable to conflict ecosystems in any region; and the representation of social science theory within the CGM, specifically tracing the implementation of the Theory of Planned Behavior. The work builds upon the experience and insight gained in the Agent-Based Simulation (ABS) Verification, Validation, and Accreditation (VV&A) Framework Study sponsored by the Marine Corps Combat Development Center (MCCDC).

## ABOUT THE AUTHORS

**Ms. Lisa Jean Bair** (formerly **Moya**) is Chief Scientist of WernerAnderson and leads its modeling and simulation (M&S) research and development effort. She is co-chair of the M&S Congressional Caucus Standing Committee in support of M&S Professional Development and Education and serves on the Certified Modeling and Simulation Professional (CMSP) Board of Directors. Ms Bair's areas of expertise include Multiple Objective Decision Analysis (MODA), Multi-Attribute Utility Theory (MAUT), Modeling and Simulation, simulation validation, human behavior modeling, and the development and application of analysis. Ms Bair has experience in analyses of alternatives; requirements evaluation; and M&S planning, use, and validation for the department of defense. Ms Bair is a PhD candidate in the Modeling and Simulation program at Old Dominion University and received an M.S. in Operations Research from The College of William and Mary and a B.S. in Applied Mathematics from Old Dominion University. Her research interests include agent based simulation, simulation theory and formalisms, human behavior modeling, and validation.

**Dr. Eric W. Weisel** is CEO of WernerAnderson, Inc. Dr. Weisel received the Ph.D. in Modeling and Simulation from Old Dominion University in 2004, the M.S. in Operations Research from the Florida Institute of Technology in 1995, and the B.S. in Mathematics from the United States Naval Academy in 1988. Prior to founding the company, he served as a U.S. Navy submarine officer on Los Angeles class attack submarines and various Navy and joint staffs with experience in nuclear engineering; navigation; and submarine, battle group and joint operations.

**CPT. Richard F. Brown** serves in the US Army as an Operations Analyst in TRAC-MTRY and is leading the research effort described in this paper.

# A Risk-Based Validation Approach for Irregular Warfare Models

**Lisa Jean Bair, Eric W. Weisel**
**WernerAnderson, Inc.**
**Gloucester, VA**
**lbair@werneranderson.com, eweisel@werneranderson.com**

**Richard Brown**
**US Army TRADOC Analysis Center – Monterey**
**Monterey, CA**
**richard.f.brown1@us.army.mil**

## CULTURAL GEOGRAPHY MODEL

The Cultural Geography (CG) agent-based model (ABM) provides a re-usable framework for representing populations within a geographic area of operations. The model is grounded in emerging military doctrine and social science theory and provides a landscape of potential futures with respect to the population's response to the actions and interactions of friendly, threat, and various other actors. The population responses to these actions and interactions and their subsequent responses are captured through measuring each individual's perception of the adequacy of Security, Governance, and Infrastructure and how their perceptions change over time. The CG modeling framework can support any training or analytical effort that requires information regarding the indigenous population.

### Objectives of the project

This work has three main objectives: (1) assess the CG Model Conflict Eco-System, use of the Theory of Planned Behavior, and Social Network Representation; (2) develop methods for assessing the risk of use; and (3) develop and implement validation plans for the overarching architecture and conceptual model implementation in general and for the use of the CG Model in particular. That is, the overarching objective is to assess and quantify, if possible, the operational utility of the CG model and provide any suggestions, rules, or limitations for using the model for the problems and uses of interest (i.e., traditional wargame applications and exploration of potential futures). Some considerations include perishibility of data, access and qualifications of needed subject matter experts, and computational complexity.

## VALIDATION OVERVIEW

Standard approaches to validation fall short of achieving the desired results for the complex behaviors generated by human behavior models and ABSs (Moya and Weisel 2007). While results validation and face validation are often used methods, the difficulties with this approach for simulations having sensitivity to initial conditions, or chaotic/emergent effects, and the difficulties with validating human based representation models is well known (Harmon 2002). In previous work, we discussed the various definitions of validation and how these definitions apply to the validation of ABS (Moya and Youngblood 2007). *Comparison* is the common thread in all of the validation definitions reviewed. In the case of modeling and simulation (M&S), the thing being represented (the model or simulation) is compared to the thing it represents. Validity is a qualitative and not an absolute assessment. A model is not valid in all cases for all pursuits; a model may be valid in some instances, but not in others. Validation can be seen as a communication activity that enables a model or simulation to meet consumer requirements to support a specific intended application and provides information to decision-makers allowing them to make effective decisions about using a model. This concept is captured by "intended use" in the DoD definition (Department of Defense 2007). Intuitively, a model's validity is based on how closely the simulation behaviors match the real world. Determining validity falls involves more scientific method than proof; we determine a model is valid when there is insufficient evidence to determine the model is invalid. That is, we have failed to reject the null hypothesis that the model is valid.

## VALIDATION OF AGENT-BASED SYSTEMS

### Agent-based systems

Although discrete event and time-stepped models dominate military simulations, multi-agent models providing human behavior representations are of growing interest to the military community (Cares 2002). The science of systems composed of human entities and their interactions lies in sociology, psychology, and military science. These systems differ from those in the "hard" sciences in that the underlying mathematics of the system are unknown (i.e., the analytic model describing the system's behavior is not known). Rather, they are described by relationships, expected outcomes under given stimuli, or theory. These systems are inherently complex and their simulations reflect that complexity. The hope is that the simple rule structures of agent based simulations (ABSs) will exhibit emergent behavior that can capture the complexities of the battle space environment.

The ABS paradigm lends itself well to modeling environments where the physical, analytical, or mathematical representation is unknown at the macro-level, such as found in social systems or when the interactions between local elements are of primary interest (Moya and Weisel 2007). The simulations are built from the bottom-up on local descriptions of behavior with embedded agent-heterogeneity, the interactions between which generate complex, aggregate behavior. That is, micro-level modeling leads to macro-level effects through interactions that occur and are governed at the local level. Thus, the behavior emerges rather than being scripted and determined in a top-down fashion.

The systems modeled using ABS may be difficult to study because of their complexity, low occurrence, and unavailability of data, or other reasons. The data available for building these models are frequently qualitative in nature and without an analytical foundation. To validate an ABS, it is important to understand what characterizes the ABS modeling paradigm from other paradigms and how these simulations may differ from other types of simulations with respect to validation.

Intelligent software agents, from which these simulation systems are comprised, are a natural metaphor for these types of complex systems (Moya and Tolk 2007). In general, they are able to act autonomously, manipulate their environments to accomplish tasks, and adapt to changes in their environments. Every agent within a multi-agent system includes an internal state representation, a knowledge base including a representation of the simulated environment, and a behavior engine that takes inputs and chooses behaviors based on the agent's state and information found in the knowledge base. The reasoning or decision-making architecture found in the behavior engine includes the agent's capabilities to react to changes in its environment and memory capabilities, as well as beliefs and goals.

ABSs are the result of the composition and interaction of many agents within the simulation environment (Hare and Deadman 2004). While this aggregation can lead to predictable results in some domains (e.g., predator-prey models, traffic models, and economic models), the mechanism leading to these results is unknown. Agents may be homogeneous or heterogeneous within their environment, have varying levels of reactivity, goal orientation, and learning capabilities, and have behavior and characteristics described using many rules and parameters. Parameter levels multiply creating a large number of possible starting states for each agent (multiplied again by the number of agents within the simulation). This creates a large number of possible trajectories for the simulation. In addition, for many of these simulations, small changes in starting parameters can result in large differences in the trajectory taken by the simulation. Further, since these interactions occur at the micro-level, these trajectories are "emergent" in that they are unanticipated from an examination of the rule sets and knowledge base governing the agents' behaviors. In addition, these models often have hidden, unobservable behaviors.

An agent is someone authorized to act for another. Agents possess the characteristics of delegacy (authority to act autonomously on behalf of the client), competency (capability to effectively manipulate the problem domain environment to accomplish the prerequisite tasks) and amenability (ability to adapt behavior to optimize performance). A software agent is an artificial agent, which operates in a software environment. An intelligent software agent pursues of the goals of its clients. As such, intelligent software agents (artificial agents, operating in a software environment, that pursue the goals of their client(s)) are a natural metaphor for many complex systems to which computer solutions are sought. (Moya and Tolk 2008) Agents are applicable to systems with the factors shown in Table 1 (Jennings and Wooldridge, 1998).

**Table 1**. Factors indicating an agent system

| **A complex environment, which may be open, dynamic, or uncertain** |
|---|
| Agents are a metaphor in the system |
| Data, expertise, or control is distributed |
| Interaction with legacy systems is necessary |

A multi-agent system (MAS) is a system that consists of multiple autonomous entities that build a population of agents. Agents carry with them a decision-making architecture and ability to perceive and react to its environment as well as potentially goals and beliefs. It is these characteristics that lead to the metaphor that makes an agent based paradigm desirable. An agent's decision-making capabilities reflects its reasoning methodology such as deliberative, tropistic, or a hybrid. Thus, neural network based, rule based, and logic based decision architectures all provide a basis for the agent based paradigm. The CG Model uses an event-based framework for next action selection and a neural network to process the reaction to the results of that event. The data supporting these reactions is developed using the Narrative Paradigm.

Perception refers to the agent's ability to sense its environment, the accuracy of that perception, and its memory of its own and others' actions. This perception feeds its beliefs of the environment. It is characterized by the ability of the agent to access the activity of the environment, the accuracy of that perception, and its memory about past events. The CG Model has no memory of previous events in the version studied nor does it address accuracy of the perception of the success or failure of events. However, access is explicitly modeled through the concept of homophily.

Some agent architectures have implicit goals while others explicit goals for which they solve the simulation system to achieve. Agent beliefs include what the agent believes about its environment, itself, and other agents within the agent system such as its beliefs about what it has perceived or sensed in its environment, the implications of these perceptions, the results of any actions it may take, the actions others may take, or the results of other agents' actions. The CG Model captures this through the social network within its Conflict Eco-System with data developed using the Narrative Paradigm. Goals, however, are implicit once the data has been developed.

**Validation challenges**

Complex systems, emergent behavior, and rapid update cycles make ABS validation particularly challenging. Both the validity of the individual agents and the validity of their interaction within their environment must be examined. Other considerations are important as well: identifying and obtaining data, especially for non-quantitative data elements; determining agent and parameter relationships and effects quantified in the referent; important dynamic elements and sensitivity analysis, especially when non-linearity is present; effects of agent heterogeneity; and the match between the system of interest and the computer instantiation.

In general, validation activities focus on two key products of the M&S development lifecycle: the conceptual model and the M&S results. Three main areas are important for validating an ABS: the conceptual model (which includes the theoretical and the mathematical models), the knowledge base, and the simulation results. The conceptual model and knowledge base validation reflects the need for a valid micro-level representation on which these models are predicated. Results validation ensures the interactions that occur at the local levels create an appropriately accurate macro-level representation of the system being modeled (for the intended use).

Physics-based modeling uses empirical data as the real world referent against which simulation results are compared. The analytic, mathematical model that forms the basis for the computation in physics-based simulations forms the conceptual model for these simulations. Validation of this conceptual model is axiomatic given broad-based acceptance of physics-based modeling in various scientific communities. When modeling other types of systems, such as social systems, the conceptual model on which to base the computation is not nearly as codified as in physics. These systems may have little or no mathematical or analytical foundation. Instead, these systems may have a theory of behavior as its conceptual referent, an idealized view of system behavior based on identified circumstances or system characteristics, or a set(s) of observed data. These form the system conceptualization against which the developed mathematical model (to be coded in the computational model) is compared to determine whether it is sufficiently complete at the desired fidelity. For ABS, the expectation is that if all of the appropriate system elements are included, the simulation's computations will result in the needed accuracy for the intended use.

To model social systems, a single model may use multiple sources to build the conceptual framework, such as subject matter expert opinion, theory, and data. There could be multiple, competing theories. Each of these sources could individually be a referent for a simulation model or could be compiled together to form a consolidated referent for another simulation model. These referents form the conceptualization of the system of interest and give the basis against which the conceptual model is compared.

The theoretical model, containing all of the agent behaviors (in an ABS), relationships, and expected outcomes, then forms the referent for the mathematical instantiation of the model used to capture the desired system model behaviors. The mathematical description forms the referent for the computational, algorithmic instantiation of the model. Each of these relationships is one-to-many; that is, there are many mathematical descriptions for a given conceptual model and there are many algorithms to capture a mathematical description. Whether the chosen capture method is appropriate depends on the required parameters and the level of accuracy desired. Lastly, expected model results and behaviors found in the conceptual model, as well as expected results from real world data, other models, direct calculation, or other sources, form the referent against which results from simulation runs are compared. Three of the main models are the theoretical model, the mathematical model, and the instantiated or coded model. The theoretical model describes the basic theory used to substantiate the description of the system under consideration. The validation process in this domain requires a determination that the system conceptualization matches available data and theory for the system containing all of the relevant aspects of that system necessary to support the intended use. When data or theory (possibly multiple theories) is in conflict, then the choices made and their justifications are documented to ensure traceability throughout the process. The theoretical model, when built, needs to drive to a sufficiently detailed system description to enable building a mathematical system description. The mathematical model is the translation of the theoretical model in mathematical language (e.g., equations) so that it can be further translated into code.

In physics-based modeling, the mathematical model is the analytic model. In the ABS construct, the mathematical model is the specific choice of how rules are implemented within the system model. For instance, the conceptual model may have a state value increasing with certain stimuli. The mathematical model is the formulae used to calculate the state value changes as a function of the relevant stimuli. More than one formulation could be possible. The coded model is the algorithm instantiated in the computer used to calculate the mathematical formulation of the theoretical system model.

Unlike in physics-based models, when an ABS is applied the theory underlying the system often is unknown. Further, data supporting model development is sparse. Therefore, there is little data available for simulation and model developers to use to build and test their models. This means that the referent against which the simulation model is compared may be difficult to obtain and, perhaps, may have to be built during the model development process. Documentation of the assumptions, references, and justifications for the choices made to develop the conceptual model can support validation, as well as support model use by providing a communication mechanism between developers, users, and decision-makers.

Any ABS validation methodology needs to address parameter interdependencies, property interactions, and behavior change sensitivities (i.e., the complex behavior space). Thus, the basis of any methodology is a firm grounding in theory. Additionally, the methodology needs to communicate necessary information about the model so effective decisions about its use can be made. Finally, the methodology needs the flexibility to meet various development levels, applications, and intended uses.

The intended use of a simulation determines the level of accuracy required. A representative simulation that accurately represents trends within the modeled system may be all that is required for some analytical applications. Accuracy improvements may be required to support experiments and training. Continuous improvement would support system predictions in applications, such as test and evaluation, where the system in the model should be truly representative of the system being modeled.

Predictive capability, and hence required accuracy, is different for ABS than for other conventional and physics-based simulations. In usage terms, ABS are probably best suited to explore the system and test hypotheses rather than for predictive analysis. In a predictive vein, the best result an ABS is likely to achieve is identification of potential trends within the system and determination of robust solutions. Such considerations suggest the need for the development of an analytic paradigm that will allow ABS to realize their potential as valid analytic tools. One type of analysis could be to experiment with social systems in ways not feasible with the real world system in order to test hypotheses about those systems. The intended use for this type of simulation might be for building intuition.

ABSs require validation on multiple levels. Since the conceptual models are not fixed within their communities the way that physical models are, the referent from which the conceptual model is derived may require some level of validation. This validation effort may consist of reconciling different theories or SME opinions. At the very least, it should consist of documenting the choices made and the support for those decisions.

The conceptual model from which the computer instantiation is derived also requires validation. That is, since the verbal description of each conceptual model can have many mathematical interpretations, the mathematical instantiation chosen requires validation. This includes but may not be limited to the agents' rule base, their interaction mechanisms, their available behaviors, and the knowledge base that drives behavior selection. One might argue that this is verification versus validation, but since this is an issue of system description system rather than of coding which lies in the realm of computer systems engineering, this assessment falls more appropriately into the realm of validation. For instance, the desired relationships and their descriptive mechanisms may be appropriately coded, but if the resulting interactions and behaviors could still result in modeling the system incorrectly (i.e., the wrong thing was modeled) since the specific mathematics of these interactions may not be known, leaving the model develop to choose (perhaps in concert with SMEs). Since the implications of choices made in bottom-up development on the emergent behaviors may not be obvious and may only be discovered during the validation process, the process of discovering this error could require a change in the

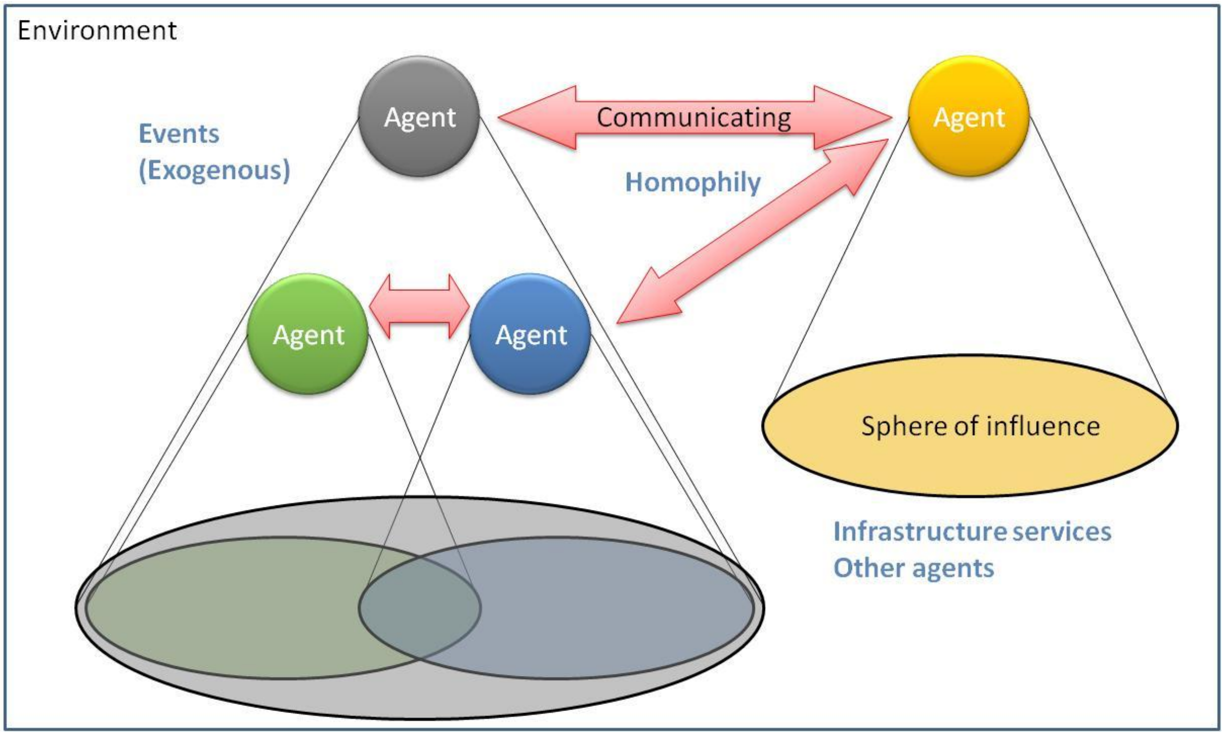


**Figure 1. CG Model as a typical multi-agent system**

conceptual/mathematical model for the system and further validation.

**CG Model as a multi-agent system**

While the CG model is an agent based simulation in the sense that agents form a metaphor for the system and the modeled environment is dynamic and uncertain [Jennings and Wooldridge 1998], the CG model follows an event-based modeling paradigm as an activation mechanism for its behavior engine. The desired output from the model following each event is the updated distribution on issue stances. The CG model uses the Narrative Paradigm as its foundational social science concept for the data generation. The Narrative Paradigm suggests that for each agent, new issue stances are a function of current issue stances, current beliefs, and knowledge of an event. It provides a basis for the strength of foundational beliefs. The Theory of Planned Behavior provides a framework for guiding agents' beliefs and intentions control mechanisms on behavior, while Homophily determines communication through the mechanism that with similar social characteristics (social status or values) tend to associate with one another. It provides a mechanism for both the knowledge of an endogenous event and direction of influence. Figure 1 is a conceptual representation of the CG Model as a typical multi-agent system.

*Events* in the model are determined through actions taken by other actors (2) such as Host Nation, Insurgents, or Control Forces (exogenous events) or resulting from the ability of agents to obtain commodities (3) through access to infrastructure (endogenous events). Actions taken by the Other Actors are based in the Theory of Planned Behavior and can be provided within the CG Model through an event list, human-in-the-loop wargaming, or other simulations. Infrastructure represents providing of goods and services and are modeled by a multi-server queues.

An agent can experience an event directly or can learn about an event from another agent. Agents know of exogenous events based on their proximity to the location of the event, and it is assumed that all agents know immediately about exogenous events that occur within their location. Agents form intentions to pass on information on events (both endogenous and exogenous) to other agents through the social network, which represents relationships and influence based on the qualitative social theories of Homophily, Influence, and Trust. The likelihood of an agent changing its viewpoint on an issue is based the homophily between the agents. The CG models uses its implementation of homophily to determine the impact of the message in a way such that "the impact of a message sent through the social network is similar to first-hand knowledge" with decreasing impact for dissimilar agents. A more detailed description of the CG model can be found in Alt et al [2009].

**Agent Taxonomy within the CG Model**

The taxonomy for agents and multi-agent systems developed by Moya and Tolk [2007], gives a properties based method for describing an agent based simulation with respect to the simulation environment, the community of agents (i.e., the multi-agent system), and the agents within the system itself. Creating a taxonomical description of agent based simulation allows the focusing of assessments to areas clearly identified in affecting agent behavior. A description of the CG model as a multi-agent system is given in Table 1with the elements of the individual agent description as shown in Figure 2. As discussed in the table, individual agents in the system have no memory except that which exists organically in the description of the Bayesian network. This lack of memory may be a limitation of the CG model depending on the specific implementation of the Bayesian network for a given application.

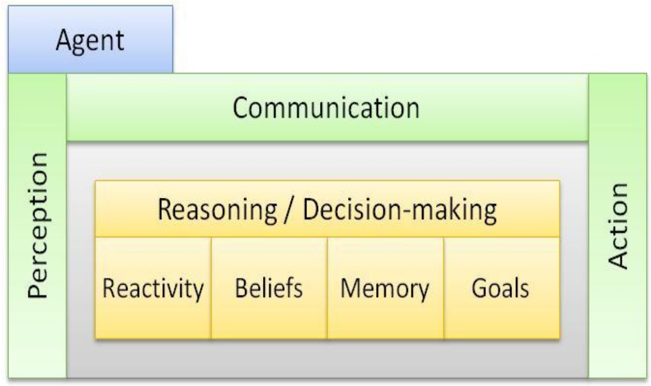


**Figure 2. Typical agent [Moya and Tolk 2007]**

The key in assessing the CG model is in assessing the characterization of the agents within the model and the methodology used to update the agents' state (current beliefs, memory and goals) and the agents' method for selecting actions based on its communication with other agents (Homophily) and perception of the environment.

**Table 2. CG model as a multi-agent system**

| Category | Classification and Description |
|---|---|
| Situated environment | Closed (outside of a federated environment): changes to the environment come from within the simulation<br>Static: agents or model inputs to the simulation cause all environmental changes<br>Deterministic: agent actions have the same effects each time with identical seeds for stochastic model elements |
| Agent population | Heterogeneous: instantiations vary between agents with homogeneous descriptor parameters (e.g., *Stereotype*)<br>Independent (between agent population vice within agent population)<br>No (community) goals |
| Perception | Partial |
| Communication | Through homophily with agents having first hand-knowledge |
| Memory | None |
| Decision making | Tropistic: Agents execute no deliberative planning functions; determine their prevailing environmental beliefs (i.e., occurrence of an event, effect of event on viewpoint on others like them) through their social network; and have only implicit goals as described by their *Beliefs*, *Viewpoints* and action selection. |

## RISK

The oft quoted "All models are wrong, some models are useful," attributed to Box, highlights that it is important to understand the risk in using a model and its simulation (M&S) results when making a decision. This enables judicious application and use of M&S. The primary purpose, and importance, of conducting validation activities is to assess the risk of using an M&S for a specific application of use. The validation process culminates in the communication of that risk to model and simulation users and the recipients of their data. This includes determining that the simulation is correct and meets requirements through software engineering and other processes but is not limited to that. It also includes providing users with sufficient information to determine if the simulation can meet their needs as well as determining the simulation's capabilities, limitations, and performance relative to the real-world objects it simulates.

Currently there is extensive work ongoing in Risk Based Verification, Validation, and Accreditation (VV&A) for the

Acquisition M&S VV&A Sub-group (Youngblood 2010). This work conducted an extensive literature survey to develop an ontology of more than 200 methods for risk based assessment methods for VV&A. Risk areas within the methodology supported by this ontology rely on the identification by the user of the role of M&S (e.g., the decisions to be supported and how the M&S supports those decisions), the importance of M&S in that role, and the validation maturity required. Thus, the risk of using a model is a function of consequence, error, and the validation process. Necessary to determining error is an understanding of the essential elements to the problem being addressed by M&S and an assessment on the degree to and manner in which these essential elements of the requisite decision problem are included in the model. These essential elements are included in the conceptual model of the system for the M&S. Missing, incomplete, or poorly described elements may indicate a higher risk level. Consequence is a direct function of the use. Error is a function of both the accuracy of the model (input data and description) and simulation results. The RBA methodology requires the identification of consequence either explicitly as an estimate of the consequences associated with an intended use along with an estimate of the probabilities of simulation limitations leading to consequences or implicitly through an importance level assignment associated with the consequences of the intended use. These methods reflect the guidance for risk assessments found in DoD MIL-STD-882D.

Therefore, the intended use of the simulation results determines the level of acceptable risk allowable in the M&S. Thus, it drives the level of representation fidelity needed and degree of acceptable abstraction in the conceptual model as well as the level and kind of accuracy, or type of validity relation, describing the sufficiency of match to the referent for an M&S. For example, some analytical applications may only need a representative simulation that accurately represents trends within the modeled system. Improvements in representation and accuracy may be required to support experiments and training. Predictive applications, such as test and evaluation, require further improvements in representation and accuracy. Therefore, intended use specifies the validation criteria for the M&S along with focusing validation efforts and providing guidelines for mitigation for missed criteria.

While risk assessment is a critical, if not *the* critical, outcome in any validation process, the complexity of the results space limits the ability to test a model's ability to address an intended use, especially in Human Behavior Representation (HBR) models. That is, even if a substantial amount of data were available for accuracy comparisons, in general the number of feasible parameter settings exceeds even a reasonable exploration of the possibilities (Moya, McKenzie, and Nguyen 2008). Stochastic models exacerbate this sampling problem. Limitations in the techniques available for validity comparisons worsen the problem further, especially in determining what to compare to in HBR models. This, then, leads to qualitative assessments of risk with little or no analytic underpinning; void rigorous, traceable, repeatable assessments; and ill-defined and poorly understood consequences both with respect to the model being used and the decisions based on its use, particularly for models in which there is little empirical data and accepted computational representation. While the RBA methodology provides an overarching process for risk-based validation assessments and a library of techniques from which to choose (i.e., it gives the possibility of *from what*), it provides little basis for *how* to choose or to apply these techniques.

Therefore, this project operated from the viewpoint of developing a validation methodology tailored specifically to the CG Model identifying potential risk areas, tests, and criteria for the spectrum of intended uses of the model. This resulted in a validation plan for the CG Model tailorable to a specific use based on guidelines for developing validation criteria embedded within the plan itself. The intent of this general CG Model validation plan was not to develop a wholesale validation of the CG model rather it specifically recognizes that each use of a model is unique by providing a basis for validating the model when used and to support ongoing development.

**Components of Risk**

The DoD Risk Management Guide for DoD Acquisition (Defense Acquisition University 2003) identifies two components toward risk in general:

1) The probability or likelihood of achieving (not achieving) a given outcome

2) The consequences of achieving (not achieving) a given outcome

There is higher risk with a higher likelihood or with significant consequences. Risk assessment includes both the identification of risk (determination of outcomes) and the analysis of risk (determination of probability and consequence of an outcome). It is in this latter aspect that M&S often plays a role. That is, the intended use for an M&S is to identify and help to mitigate risk, identified as part of some specified objective. However, the use of M&S

in this analysis poses an inherent source of risk. The sources of risk could lie in the development of the model, development risk, or in the running of the simulation, operational risk (Modeling and Simulation Coordination Office 2004b). Development risk is that the model does not meet the requirements for its intended use. Operational risk is that the M&S exhibits insufficient accuracy to provided needed information. The V&V process addresses both these risk areas. When considering intended use, risk can be described generally using the three familiar error types:

1) Type I Error: Reject correct information; the information provided by the M&S is not used in solving the problem even though the information provided is correct.

2) Type II Error: Accept incorrect information; the information provided by the M&S is used in solving the problem, however, the information provided is incorrect.

3) Type III Error: Solve the wrong problem; the information provided by the M&S is irrelevant to the actual problem to be solved.

Validation primarily assesses the Type II error. The *Verification, Validation, and Accreditation Recommended Practices Guide* (*VV&A RPG)* discusses this as follows (Modeling and Simulation Coordination Office 2004a). When assessing the consequences of using incorrect data in a decision, considerations include who is affected, the severity of the effect, and the visibility of the consequences. Development risk assesses the effect of not meeting requirements, the likelihood of a deficiency, and the probability that a deficiency will cause the M&S not to meet requirements. These assessments drive toward the fundamental assessment of whether the M&S support the intended use. Operational risk assesses the probability of making an incorrect decision, the effect and visibility of making an incorrect decision, and specific user considerations.

When deciding to use information from an M&S, quantitative assessments might be provided in the following ways. First, risk might be assessed as *Risk(Outcome)=Pr(Outcome)×Value(Outcome)*. Alternatively, a region of risk acceptability might be identified as shown in Figure 3, adapted from (Guarro and Vessely 2004).

This leads to a specific, yet general, methodology for risk based validation assessments. Consequence of using the model is determined by the user not the validator.

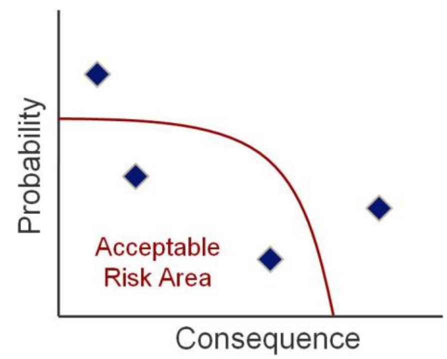


**Figure 3. Notional Region of Risk Acceptance**

Validation processes provide an estimator for the likelihood of having the bad outcome as a result of a type II error (i.e., type II error is zero if all trajectories are tested with a positive result). Stronger validation processes and more testing reduce the error associated with this assessment, tightening error bands, and allowing for a more confident assessment of risk with the model. When no more testing can reduce the risk assessment (e.g., if no matter how tight the error bands get risk will still be assessed at YELLOW or if no matter the probability of a type II error the possible consequence has the same risk assessment), then the amount of validity testing is sufficient. This matches intuition in that if the consequence is low, then less testing is required.

While quantitative assessments of risk are desirable, these are not always possible. The *VV&A RPG* suggests that qualitative assessments of risk can be applied, noting that the DoD MIL-STD-882D: Standard Practice for System Safety provides an accepted example for these qualitative assessments. This military standard provides qualitative descriptions for both probability and consequence categories, which could be adapted to the simulation context of interest.

**Risk and Intended Use**

A key insight from the ABS VV&A Framework Study, commissioned by the MCCDC Operations Analysis Division (OAD) to address shortcomings of the *VV&A RPG* with respect to the simulations of interest (IW-ABS for analysis), is that the validation of models in support of analysis resides within the analysis process itself. That is, validation cannot be decoupled from the analysis plan, process, and results. Results from applying the developed ABS VV&A Framework Study determined that generic model descriptions and applications of use were insufficient to address the appropriateness of using an M&S. It indicated the critical importance of clearly specifying the intended use for the M&S in the

**Table 3. MIL-STD-882D example with color coding**

| SEVERITY/ PROBABILITY | Catastrophic | Critical | Marginal | Negligible |
|---|---|---|---|---|
| Frequent | 1 | 3 | 7 | 13 |
| Probable | 2 | 5 | 9 | 16 |
| Occasional | 4 | 6 | 11 | 18 |
| Remote | 8 | 10 | 14 | 19 |
| Improbable | 12 | 15 | 17 | 20 |
| RISK LEVEL | 1-5 = High | 6-9 = Serious | 10-17 = Medium | 18-20 = Low |

validation efforts beyond generic descriptions. It became obvious that validation is an analysis process, intrinsically intertwined with the analysis for which the (M&S) is employed as a tool.

Consider the mishap risk assessment values from MIL-STD-882D with the addition of color coding of risk levels (e.g., RED, ORANGE, YELLOW, and GREEN) shown in Table 3. Examples of consequences found in MIL-STD-882D relate consequences to specific outcome areas such as safety, cost, performance, schedule, political, or other areas. However, to make this effective, consequence and the probability of achieving that consequence must be clearly defined in order to craft the decision problem.

**Formalizing Risk**

A fundamental paradigm shift is required in assessing the risk of using a social science based simulation to support decision-making. In particular, to effectively assess risk in using the simulation, it is critical that consequence be clearly understood and articulated with respect to its use. This is directly in line with the research described above with one fundamental difference. The research above presumes that V&V efforts can be specified after the probability of a consequence due to simulation limitation is assessed. However, it is unclear how a simulation limitation can be assessed prior to validation activities. Validation uncovers limitations rather than applies efforts to limitations.

Therefore, rather than using a priori risk assessments to guide validation efforts (i.e., arbitrarily equating importance of a simulation element to risk), validation efforts ought to lead instead to an assessment of risk. This needs to be bounded in a clear definition of the meaning of consequence and thereby the meaning of the probability of achieving that consequence of using the model.

This implies that consequence is only important as it relates to using the output from an M&S that is "incorrect." Informally, a consequence only matters if it occurs because of information provided by an M&S. If the outcome would have occurred anyway, there is no risk in the sense of M&S. Thus, the probability must be related to the likelihood that the information provided by the M&S will incorrectly lead to an alternative solution that will cause a worse consequence than would have occurred otherwise.

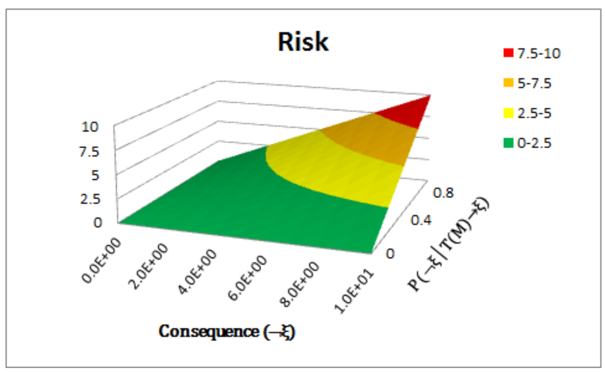


**Figure 4. Risk assessment shown as a classic s-curve**

More formally, $\beta = P(\neg\xi \mid T(M) \rightarrow \xi)$. Here $\xi$ is a state, trajectory, or other condition, perhaps formalized as a logical statement in first-order predicate calculus, Z, or some other logical language, that is predicted to exist in the simulated system as a result of simulation-based analysis (i.e. $T(M) \rightarrow \xi$ ). Likewise consequence is the "cost" of $\neg\xi$ occurring in the simulated system, perhaps in relation to $\xi$. Consequence may be measured, or estimated, on a variety of scales. Utility theory may be applied directly to the consequence axis of the risk function.

Therefore, risk assessments in the use of simulation will need to clearly articulate the possible alternative consequences of using the M&S and their relationships to each other should an alternative path be chosen as a result of using the model. Only then, can probabilities of achieving these outcomes be crafted and, potentially, estimated through a validation process.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the help and support of TRAC-MTRY in the Validation Methods for Assessing Conceptual Models and Risk in Modeling & Simulation (M&S) for Irregular Warfare (IW) Analysis project.

## REFERENCES

Bair, L. J. & Weisel. E.W. (2010) Determining operational utility of the cultural geography model. In Program Committee (Ed.) *Proceedings of the 2010 International Summer Simulation Conference*, Ottawa, CANADA: The Society for Modeling and Simulation International.

Cares, J.R. (2002). The use of agent-based models in military concept development. In E.Yücesan, A.Chen , J.L.Snowden, & J.M.Charnes (Eds.). *The 2002 Winter Simulation Conference* (pp. 935-939). Piscataway, N.J.: IEEE.

Defense Acquisition University. (2003). *Risk Management Guide for DOD Acquisition*. Ft. Belvoir, VA. Defense Acquisition University Press.

Department of Defense (2000). *MIL-STD-882D: Standard Practice for System Safety.* AREA SAFT. AMSC. Retrieved from http://safetycenter.navy.mil/instructions/osh/milstd882d.pdf

Department of Defense (2007). *Department of Defense Directitve (DoDD) 5000.59: DoD Modeling and Simulation Management*. USD(A&T).

Department of Defense. 2000. *MIL-STD-882D: Standard Practice for System Safety*. AREA SAFT. AMSC. Retrieved from http://safetycenter.navy.mil/instructions/osh/milstd882d.pdf

Guarro, S., & Vessely, W. E. (2004). Risk Modeling. In Rainey, L. B. Davis, P. K. (Eds.) Space Modeling and Simulation Pages (pp. 369-450). AIAA.

Hare, M. & P. Deadman. (2004). Further towards a taxonomy of agent-based simulation models in environmental management. *Mathematics and Computers in Simulation*. *64*, 25-40.

Harmon, S.Y., Hoffman, C.W.D., Gonzalez, A.J., Knauf, R., & Barr, V.B. (2002). Validation of human behavior representations. *Foundations for V&V in the 21st Century Workshop (Foundations '02)*. San Diego, CA: The Society for Modeling and Simulation International (SCS). (October 2002), B3-1-B3-34.

Jennings, N.R. & Wooldridge, M. (1998). Applications of intelligent agents. In N.R.Jennings & M.Wooldridge (Eds.), *Agent Technology: Foundations, Applications, and Markets* (pp. 3-28). Berlin: Springer-Verlag..

Modeling and Simulation Coordination Office. (2004a). Special topics: Risk and its impact on VV&A. *The Verification, Validation, and Accreditation Recommended Practices Guide*. Retrieved from http://vva.msco.mil/.

Modeling and Simulation Coordination Office. (2004b). *The Verification, Validation, and Accreditation Recommended Practices Guide*. Retrieved from http://vva.msco.mil/.

Moya, L. J. & Tolk, A. (2007). Towards a taxonomy of agents and multi-agent systems. *Proceedings of the 2007 Spring Simulation Conference: Agent Directed Simulation (ADS-10).* San Diego, CA: The Society for Modeling and Simulation International (SCS).

Moya, L. J. & Weisel, E. W. (2007). Towards a methodology for validating emergent behavior simulations. Program Committee (Ed.) *Proceedings of the 2007 European Simulation Interoperability Workshop*, Orlando, FL: Simulation Interoperability Standards Organization (SISO). 07E-SIW-055.

Moya, L. J., McKenzie, F. D. & Nguyen, Q.-A. H. (2008) A visualization technique to support human behavior representation model rule validation. *Simulation & Gaming: An Interdisciplinary Journal of Theory, Practice, and Research*. *39* (1), 101-117.

Moya, L.J. & Weisel, E.W. (2008). The Difficulties with validating agent based simulations of social systems. *Proceedings of the 2008 Spring Simulation Interoperability Workshop*. Ottawa, CANADA: Simulation Interoperability Standards Organization (SISO).

Moya, L.J. & Youngblood, S.A. (2007). Clarifying Validation for Agent Based Simulations. *Fall Simulation Interoperability Workshop*. Orlando, FL.: Simulation Interoperability Standards Organization (SISO).

Youngblood, S. (2010). *Risk-Based VV&A (RBA) Methodology: The Concept of M&S Use Risk*. Paper presented to at the Military Operations Research Society Symposium 2010 (MORSS 2010), Quantico, VA, 22-24 June 2010.