Title: **Validating a Computational Model of Decision-Making using Empirical Data**

Authors: Ju-Sung Lee, Carnegie Mellon University
Kathleen Carley, Carnegie Mellon University
Judith Effken, University of Arizona
email: jusung@andrew.cmu.edu

Extended Abstract

Let us be frank with ourselves and pose the question, why a computational model in the first place? To answer this, we must demonstrate that analytic and statistical methods neither adequately explain our process nor properly predict the outcomes for anticipated scenarios, while computation achieves both. Implicit in this endeavor is the requirement for us to define the criteria by which we convince ourselves, and others, that the model is both correct and useful. That is, at minimum, the model must both fit the given data and generalize to new data using parameters that are sensible, meaning that their derivation from the empirical data is clear. These tasks constitute the ever-present challenge of justifying and validating a computational model. The ideal model is sufficiently predictive, so as to obviate the need for collecting new data, and yet maximally parsimonious, so as to be easily understood and usable.

While many developed models of all flavors mathematical, statistical, and computational can describe a host of social phenomena, these phenomena are relatively simple compared to the ones that still require quantitative explication that is accurate enough to inform consequential decisions such as policy and organizational strategy. In short, the state of the art in modeling still imposes no standards on many of the aforementioned validation issues. At what point does, say, a statistical model become inadequate to capture the complexity of the problem and needs to be upgraded to a computational model? Sometimes, we are attracted to the dynamic nature of computation, which gives us a means to observe the evolution of the system. Other times, the system can only be accurately described as a dynamic process. As more and more models arrive on the scene, we need to ask if an extant model is sufficient or if we need to develop yet another, specifically tailored to our specific problem.

In this paper, we validate the predictive ability of a generalized computational model of decision-making using empirical data from two distinct studies. The first study comprises data obtained by the College of Nursing at the University of Arizona. Researchers collected longitudinal performance data of a dozen nursing units across several hospitals along with a host of other variables, descriptive of the network structure and workloads of these units [1]. The second study looks at the performance of military units operating under an A2C2 command-and-control structure. These units participated in a virtual war-game in which decision-makers performed tasks, allocated resources, and co-operated with their teammates to accomplish geographically localized military objectives (e.g. taking a beach and airport) [2]. Finally, our decision-making engine is one of the components of ORGAHEAD, a model of organizational adaptation developed at Carnegie Mellon University [3]. We will use this model to separately parameterize and fit the data from each study.

Assuming we can achieve some fit to the data with our model, the primary challenge is to ascertain the balance point between parsimony and predictive power. That is, how much of the problem do we need to simulate? Often times, this requires a trial-and-error search while constantly asking, at point does my model fail to predict? The analyses in this paper represent one trial of such a search; as you will read below, by simplifying the complexity of numerous data elements, we impose a high degree parsimony in the model specifications while testing its fit with the A2C2 and nursing study data. And throughout this search, we need to explicitly match key constructs between the model and data; these include variables of classical modeling as well as those defining the dynamic process:

1) **The dependent measure**. In the model as well as data, our primary outcome measure is decision-making accuracy: how correct are the organizational units in solving their decision tasks?

2) **The predictors**. For A2C2, we use the structure of the meta-matrix [4] to predict, both analytically and computationally, the task performance. For the nursing study, we employ education and experience levels of people and the network structures from the various nursing units as our predictors.

3) **Aligning or reducing complexity**. The mapping of empirical data to a model's process is perhaps the most crucial component of modeling. Determining appropriate complexity of the parameters, the dynamic process, and outcome measures are distinct issues. Again, the goal is to achieve that optimal balance between parsimony and predictive power. And again, these decisions are subjective and often warrant experimentation.

    a. Complexity of the **problem parameters**: In the nursing study, the difficulty, or complexity, of the decision problem was assessed using just a handful of predictors of performance, including co-morbidity in patients (i.e. how many medical issues did each patient have?) and staff evaluations of turbulence and complexity in their units. In A2C2, task dependency on resources defined the level of complexity per task. That is, the more resources a task required, the more complex we deemed the task to be.

    b. Complexity of the **dynamic interaction**: At one extreme, the model can attempt to capture each and every signal transferred between the information processing units, being people or books or computer systems. However, that level of detail is often not necessary to achieve adequate predictive power. For our validation efforts, we assume that such information is reducible and treat signals between agents in the model as a gestalt. And, since we lack detailed communication data, we can only estimate the degree of simplification.

    c. Complexity of the **outcome measures**: The outcome measure for each study and the model is a single scalar, thus keeping complexity low. For the nursing study, our dependent or outcome variable is frequency of medication errors, while in A2C2, it is task accuracy. And as mentioned earlier, in ORGAHEAD, decision-making accuracy is the appropriate analogous measure.

d. Complexity of **time**: While the nursing study occurred over a period of six months, we have no data on the rate of decision-making. Instead, we collapse the model time horizon to a duration that yields sufficient variance in the outcome measures. For the A2C2 study, we do have rates of decision-making; however, this validation exercise seeks parsimony as well as exactness, and, so we similarly collapse the time horizon.

Our results demonstrate that an analytic or statistical model is insufficient to even fit the results, let alone describe the dynamic process. For both studies, key indicators of performance did not correlate adequately with the empirical findings. In the A2C2 study, resource distribution is often a strong predictor of task performance but did not correlate with the results of the study. Our model results yielded a much better fit; refer to figures 1 and 2. Note that the relative orderings of performance matter here, rather than their actual values. In the nursing study, the variables, which correlated highly with measures of performance, still failed to yield a proper ordering of performance, while model results again proved to be a better fit.

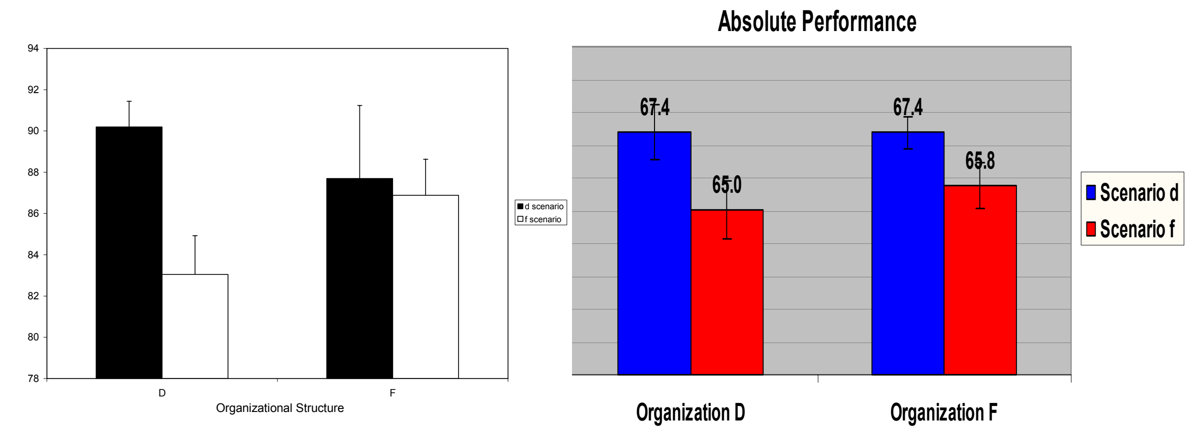

Figure 1. A2C2 Study Results (left) vs. Model Results (right).

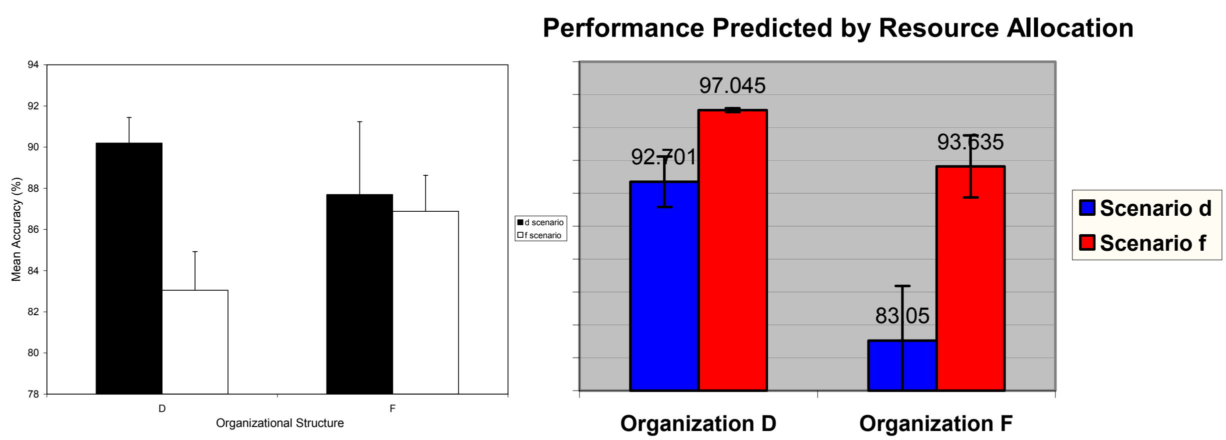

Figure 2. A2C2 Study Results (left) vs. Analytical Results (right).

These findings demonstrate that the non-linear, dynamic process through which the states of initial parameters evolve is essential in modeling these data, thereby justifying the application of the model. So, as additional evidence of validity, we will present a deconstruction of the mechanisms that allowed the model to fit better than the analytical results.

References

[1] Effken, Judith A., Barbara B. Brewer, Anita Patil, Gerri S. Lamb, Joyce A. Verran, and Kathleen M. Carley, 2003. "Using computational modeling to study the impact of workplace characteristics on patient safety." Working paper. University of Arizona.

[2] Aptima, Inc., 2002. "A2C2 Experiment 8 Data Files Guide."

[3] Carley, Kathleen M. and David M. Svoboda. 1996. "Modeling organizational adaptation as a Simulated annealing process." *Sociological Methods and Research.* 25(1): 138-168.

[4] Carley, Kathleen M., 2002. "Smart agents and organizations of the future." *The Handbook of New Media*. Edited by Leah Lievrouw & Sonia Livingstone. Chapter 12: 206-220. Thousand Oaks, CA: Sage.

[5] Carley, Kathleen M., 2003. "Validating Computational Models." *CASOS Working Paper Series*. Carnegie Mellon University.

[6] Burton, Richard and Børge Obel, 1995. "The validity of computational models in organization science: from model realism to purpose of the model." *Computational and Mathematical Organization Theory*. 1(1): 57-71.

[7] Naylor, Thomas H., J.M. Finger, James L. McKenny, William E. Schrank, and Charles C. Holt, 1967. "Verification of computer simulation models." *Management Science*. 14(2): 92-106.

[8] Kleindorfer, George B., Liam O'Neill, and Ram Ganeshan. 1998. "Validation in simulation: various positions in the philosophy of science." *Management Science*. 44(8): 1087-1099.