

# The Challenge of the Black Box: Technosocial Predictive Analytics (TPA) and the Real World

Lars Hanson and Anne Russell

Science Applications International Corporation  
4001, North Fairfax Drive, Suite 725, Arlington, VA 22203  
{Hansonlp | Russellav} @saic.com

## Abstract

The term “black box” traditionally refers to understanding systems in terms of input and output characteristics. In Technosocial Predictive Analytics, the term most often is applied to models, since their specific internal workings either are unknown or need not be considered. This paper examines the consequences of approaching the real world itself as a black box, and how some current verification and validation challenges may be simplified thereby.

## The Black Box

Technosocial Predictive Analytics (TPA) modeling discussions often lead to invocations of the “black box,” the term coined (Cauer 1941) to describe the process of using transfer functions to relate inputs and outputs in linear time-invariant systems. In TPA modeling, the term black box is used to refer to models for which the specific internal workings either are unknown or need not be considered. However, the term is used less rigorously and the systems under discussion are more complex. There are three significant factors in the real world which can affect the basis and effectiveness of technosocial modeling:

- *Responses are not necessarily linear.* The transfer functions may take any of several forms and may involve thresholds rather than graduated responses;
- *Responses are not time-invariant.* The same set of stimuli will not always produce the same set of responses over time;
- *Behavior is influenced by its predecessors and antecedents,* and by perceptions of what might happen. As such, there is always a cost associated with a given action. Furthermore, once a stimulus has been applied, one cannot restore or reset the system to its previous state.

If the TPA aim is to accurately predict the set of real-world outputs or responses given a set of inputs or stimuli,

modelers can treat the real world as a black box, and in so doing simplify validation and verification without impairing the effectiveness of a TPA system. In attempting to represent the “real world,” modelers are forced to think about constructing models which replicate ultimate real world truths. Yet the nature of ground truth is that what was true yesterday may or may not be true tomorrow. Does that mean the model was wrong if it does not discover that truth? Not necessarily. Rather it means that, if one is modeling inputs based on real world events, outputs must conform to real world results.

By treating the complexity of the real world itself as a black box, the modeler is freed from using validation techniques to “prove” beyond a measure of doubt that the TPA model is “correct.” Instead, validation can be accomplished by demonstrating and measuring the degree of congruence between predictions and actual results. As in any system, however, it is essential to define as clearly and as precisely as possible the relevant inputs and outputs.

## Modeling Challenges

The most challenging tasks in TPA lie in applying established scientific methods and theories to the highly complex and asymmetric puzzles created by human interactions. No matter what the technique – structured argumentation, statistical or agent-based modeling, system dynamics modeling, content analysis, etc. – the common hurdles are at once both simple and daunting:

- *Capturing models for computational analysis* – While mathematically based TPA models are growing in number, the majority of human event based models do not easily lend themselves to computational environments, leaving the TPA field bereft of the majority of thinking on the complex phenomena that TPA seeks to explain. Greatly complicating matters is the visceral distrust that many human event modelers have of attempts to distill humanity into binary strings. Yet, teasing the models into computational forms is a momentous and necessary task for the field to advance.

- *Definitions* – Defining a particular human event based problem clearly enough for computational modeling is difficult. Most human event based issues have multiple and frequently competing definitions, almost all of which are inexact. How does one distinguish “rebellion”, for example, from other forms of armed conflict – a coup, an insurgency, or a civil war – all of which share common dependent variables?
- *Paucity of data* – Distinct social or behavioral phenomena are a relative rarity. How does one build and test models for “coup” when there are less than 3500 instances of armed conflict worldwide from 1946 to 2006? How will one build a coup data set when coups are grouped together with wars, invasions, revolutions, insurgencies, and armed separatist or nationalist movements among other armed conflict events (Schrodt 2001)? 3500 may seem like a large number of instances, yet these occurrences are of differing types and are distributed over various geographic and cultural regions and under so many varying conditions that in fact the numbers of truly similar events are in the single digits.
- *Propositions* –The majority of political, social, and behavioral theoretical models rely on subjective interpretations of speculative propositions presumed to represent “truth.” How can one verify that a particular theoretical proposition has any connection to events in the real world? Are there sound, repeatable experimental bases for such models? Are the premises falsifiable? How does one know whether the theoretical model will act as expected? How will a model predict the unknowable until it is observable?

TPA modelers will confront these four key challenges consistently with any attempt to predict events in the real world. However, by approaching the real world as a black box, the modeler need only find a transfer function which will produce the observed set of outputs for a given set of inputs, and need not resort to any speculation about internal causes of the behavior. Rather the modeler must focus efforts on assuring that outputs reflect inputs well enough and with sufficient accuracy that multiple end users will have confidence the results are reliable, reproducible, and actionable. Such an approach allows the use of results based on reproducible experiments.

### Quasi-Experimental Design (QED)

Given the practical and ethical difficulties inherent in human event analysis, developments in the application of quasi-experimental design (QED) (Shadish, Cook, and

Campbell 2001) techniques offer great potential for interpreting existing real world data.

Under the right conditions and circumstances, QED allows human events modelers to re-use existing observational data which has been gathered for other purposes to conduct additional analyses which constitute experiments differing from the original. In conducting and analyzing such experiments, the modeler can address only the data subsets of interest, thereby significantly reducing both modeling and computational complexity. Furthermore, causal inferences and relationships can be detected and measured without first requiring a human event-based model to “explain” the relationships.

Thus, under the right conditions, the use of quasi-experiments makes it possible to approach the real world as a black box. One has examined and quantified the causal inferences to be drawn between inputs and outputs without regard to the actual internal mechanism causing the observed relationships. Furthermore, the analysis of the experimental data provides strong arguments in support of the relationships being modeled.

Finally, the use of experimental techniques such as QED allow modelers to divorce the evaluation and validation of techno-social predictive analytical performance from any need to evaluate and validate the underlying sociological, political, psychological, economic, or other human event-based theories explaining the behavior observed. While some analysts and human event modelers may find such an approach less than satisfying, having solid experimental results supporting the model’s relationships and its repeatability over time, will only bolster the credibility and usefulness of the model in real world applications.

It may well be that not every phenomenon being modeled will require a black box approach to the real world. It may also be that not all phenomena of interest can be expressed adequately based on a black box approach. It is almost certain that more complex models will require a combination of techniques to capture the phenomena of interest adequately. However, it is clear that in all cases, whatever models are used must at least pass the real world black box test – that alone is sufficient, and to fail to do so is to fail validation.

### Metrics Challenges

Often metric regimes are geared more toward satisfying transfer function metrics for computational analysts, specifically those who build and use modeling tools to explain human event based phenomena. Such regimes all too often are explained in mathematical terms and focus on the performance of the engine driving the model.

Yet by its very nature, metrics in technosocial predictive analytics must satisfy the two distinct audiences with

different needs: the analysts who must test the model's theoretical underpinnings, and the operational user who must have reason to trust that their model mimics the real world with sufficient fidelity for their needs.

Analysts and/or social scientists need metrics that allow them to measure the social science-based construct of their theoretical world, i.e., the indicators and/or combination of indicators that lead to certain complex events. For example, if the TPA modeler asserts that a history of abuse is always, frequently, or rarely a necessary condition for an escalation in violence to occur, how will the modeler prove the hypothesis? Under what conditions will the hypothesis hold true? If the hypothesis holds true, how does one treat an anomaly? When do anomalies become evidence of new truths?

When assessing TPA systems, how does one devise evaluation metrics which can meet scientific standards, satisfy analytical expectations, and provide credibility for the TPA model's average user? There are two basic metrics questions:

1. Did the modeler build the model correctly – i.e., did the modeler *verify* (DoD 1994; DoD 1995) that the model accurately implements the modeler's concept?
2. Did the modeler build the correct model – i.e. did the modeler *validate* (DoD 1994; DoD 1995) the degree to which the model is an accurate representation of the real-world from the perspective of the model's intended uses.

Of these two, verification generally is easiest to answer, since the model behavior was specified before having been written into a computer program. Validation, however, is far more difficult, highlighting why problems of model construct, definitions, paucity of data, and propositions quickly become relevant to TPA. The core question reduces to one of assessing the degree to which the model's behavior faithfully emulates and forecasts real-world behavior. As such, when validating real world events, questions of relevancy, specific core behavioral definitions, and so on are significant.

Traditional accuracy, precision, and recall (APR) metrics are based on the assumption a model's predictions are either true or false. That may or may not be the case in predicting human, individual, and group behaviors. If a model assigns probabilities to different possible outcomes, then clearly traditional APR metrics are inappropriate. Furthermore, once an outcome occurs, all real-world probabilities collapse to certainties. It is very difficult to obtain reliable estimates of how likely different possible outcomes actually were. "Ground truth," in this case, is elusive at best, and most likely not accessible. For this reason, traditional classifier metrics are insufficient of themselves, most particularly in measuring TPA accuracy.

Additional metrics must be developed for comparing model performance against ground truth.

By approaching the real world as a black box, the TPA system evaluator is relieved of the burden of having to demonstrate that an asserted internal functioning is in fact correct. The social or behavioral reasoning behind any specific transfer function employed is irrelevant, *as long as the transfer function faithfully relates inputs and outputs in the real world.* This greatly simplifies comparison between a TPA model's results and real-world events ("ground truth"), as it is not necessary to confirm the accuracy of any particular theory, only to validate that the transfer functions relating inputs to outputs are accurate.

Systems engineers tend to consider a model as a series of algorithms which process data to produce an output. Social scientists, on the other hand, want models to test some particular social or behavioral theory, to explain phenomena, and to implement and test assumptions about the real world. However, in TPA a specific model need not embody any particular theoretical social science construct, but need only implement transfer functions which accurately relate real-world inputs and outputs.

Operational end users tend to think of models as a means to "explain" human-based phenomena in terms of cause and effect, and to forecast events and the potential consequences of actions. Why a particular model is effective in explaining an event and the world context within which the model makes a prediction are less important than the fact that it does so successfully. Because operational users want to know what factors drive the phenomena – e.g., what factors would cause a coup to happen or not happen at the predicted time – validation is essential. It is the process which confirms the model's relationships between inputs and outputs. Without such testing, the operational end user has no basis for trusting the model.

## Conclusions

The success of any TPA modeling system ultimately rests on the confidence all parties have in the validity, reliability, and utility of the system's output predictions. Demonstrating a model's credibility requires a defensible, verifiable, and repeatable approach to measuring the effectiveness and accuracy of the system. While a necessary component of any evaluation effort, traditional metrics cannot accurately assess a TPA system.

Treating not only the TPA models but also the real world as black boxes reduces validation to comparing the responses of the two systems to given inputs. This also permits validating a model against experimental behavioral science results. Such comparisons are not easily done experimentally – while one can run the model several

times and measure its behavior, it is not possible to run the real world several times for comparison. In that case, the question of user confidence ultimately rests on the system's performance over time.

## References

Cauer, W. 1941. *Theorie der linearen Wechselstromschaltungen*, Vol.I. Akad. Verlags-Gesellschaft Becker und Erler, Leipzig.

DoD Directive 5000.59. 1994. "DoD Modeling and Simulation (M&S) Management."

DoD 5000.59-P. 1995. "Modeling and Simulation Master Plan."

Schrodt, P. A. 2001. "Automated Coding of International Events using Sparse Parsing Techniques," Paper presented at the International Studies Association, Chicago, 21-24.

Shadish, Cook, and Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, 2<sup>nd</sup> Ed. New York: Houghton Mifflin Company.