A Guide to Analytic Techniques for Nuclear Strategy Analysis

Prepared for:

JS/J-3/DDGO
STRATCOM J9
OSD/ASD (R&E)

by:

Belinda Bragg and George Popp
NSI, Inc.

For Further information about this document please contact Belinda Bragg bbragg@nsiteam.com

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Concepts & Analysis of Nuclear Strategy (CANS) project undertaken for US Strategic Command (USSTRATCOM) was tasked to examine the utility of alternative analytic techniques for assessing nuclear force attributes and sufficiency under a variety of changed conditions. The CANS software developed during this effort is designed to enhance the planning process by guiding the analyst through the process of selecting appropriate alternate analytic techniques.

Using the CANS software[1], an analyst first starts with a question. They then select the appropriate level of analysis and the availability and type of data for answering that question. The application highlights the analytic techniques that may be used to help answer the question. The analyst can read a brief description of the technique and the resources are required to implement a research design using that technique. The analyst may also read an in-depth write up of how to use the technique in the context of nuclear deterrence, and in some instances examples of the application of the technique to the CANS problem space. This guide is a compilation of all the write-ups produced for the CANS software. These write-ups fall into two categories:

PART 1: Generic technique write-ups

These focus on a particular analytic technique. These provide the reader with a thorough introduction to a specific analytic technique. They are intended to provide enough information to enable the reader to determine the utility and practicality of that technique to problem they wish to examine. They are not intended as a guide for the application of that technique.

At the end of each write-up is a requirements section that discusses the data, time, tools, cost, skill set and expertise required to implement such a technique. For the purposes of comparison a coding scheme (outlined below) was developed to provide user with a way of comparing between different techniques.

PART 2: Examples of the application of modeling techniques to the CANS problem space

As part of the CANS effort various modeling and analysis projects were undertaken. These were designed to both demonstrate the utility of a specific technique to the nuclear strategy context and provide further insight into relevant questions. As well as providing detailed reports of their efforts, contributors to the CANS modeling effort also provided brief write-ups of their work for the CANS Software.

[1] The CANS Software is available upon request from Tom Allen tallen@nsiteam.com

REQUIREMENTS SECTION CODING SPECIFICATIONS

DATA

Category	Score	Definition	Example
Very Low	1	Qualitative data can be acquired by the user using a readily available source in a short period of time.	CIA World Factbook online - country level information on GDP, population, government type used qualitatively (in narrative form)
Low	2	Qualitative data can be generated by user from readily available sources in short amount of time. Qualitative data can be complied from a number of readily available sources	Take CIA world factbook data and convert to a quantitative data set of country characteristics (GDP, population, government) Take information from CIA World Factbook, newspapers, policy publications to create a written background report on a specific country
Average	3	Quantitative data can be generated by user from existing, multiple sources but must be complied and may require recoding. Qualitative data requires use of extensive existing sources as well as SME input	Take data country level data from CIA World Factbook, World Bank and Militarized Interstate Dispute Dataset (MID) to create a dataset that combines trade and foreign investment data and demographic information about a country with data about their involvement in interstate disputes. Synthesize information from newspapers, policy, and academic publications to produce an in-depth report on a single country or background reports for a number of different countries
High	4	Quantitative data must be collected by the user for a small range of variables across less than 10 cases or fewer than 10 years for one case Qualitative data requires use of extensive existing sources, SME input and foreign language materials across a number of cases	Implementation of a public opinion survey regarding support for US nuclear policy (less than 20 questions) in all NATO countries. Synthesize information from newspapers, policy, academic publications, SME and decision maker interviews to produce an in-depth report regarding each NATO member government's likely response to development of an ABM by the US.
Very High	5	Quantitative data must be collected by the user for a large range of variables across more than 10 cases, one case for more than 10 years or multiple cases across time. Qualitative data requires use of extensive existing sources, SME input and foreign language materials across a number of cases over time	Implementation of a public opinion survey regarding changes in support for US nuclear policy (more than 20 questions) in all NATO countries, over a two year period (data collected monthly) Synthesize information from newspapers, policy, and academic publications to produce an in-depth report on changes in UN member government's responses to US ABM plans.

SET-UP TIME

Time required to collect and prepare data, and develop model

Category	Score	Coding
Very Short	1	Less than 1 month
Short	2	Less than 4 months
Average	3	Less than 9 months
Long	4	Less than 15 months
Very Long	5	More than 15 months

EXECUTION TIME

Time required to run model, analyze results, refine model if needed

Category		Definition
Short	1	Less than 1 day to run model, analyze results, refine model if needed
Average	2	Less than 1 week to run model, analyze results, refine model if needed
Long	3	Less than 1 month to run model, analyze results, refine model if needed

TOOLS

Category	Score	Definition
none	0	No specific tools are required for this technique
minimal	0	Implementing this technique will require the use of basic word processing and spreadsheet software capabilities, such as Word and Excel
average	1	This technique requires access to a full statistical software package such as SPSS or STATA.
specialized	2	This technique requires access to a range of software packages, and / or specially designed software programs and models for analysis and presentation of results

COST

Cumulative score on other requirements

Category	Score	Definition
$	3.25-8 (lowest score=5.25)	Score based on how much technique requires on other requirements
$$	9-11	* Code score on each requirement (data, time, skill) from lowest to highest.
$$$	12-14	* For each technique add scores to get raw cost (range = 3.25 - 17)
$$$$	14-17	* Collapse to create 4 categories

SKILL SET / EXPERTISE

Category	Score	Definition	Example
Specialized / Minimal	0	Enough familiarity with a specific technique and / or software application to be able to implement an existing model and answer a pre-defined research question	Ability to run basic statistical analysis in SPSS or another statistics package and provide a simple interpretation of results if the model is specified.
Specialized / Average	1	Sufficient experience and training with a specific technique and / or software application. Ability to take a pre-defined research question, build and run an appropriate model	Ability to construct and complete a subjective decision matrix and analyse and discuss the results in light of a pre-defined research question
Specialized / Advanced	2	Advanced training and considerable experience with specific technique and / or software, ability to program software as needed Ability to define appropriate research question, build and run a matching model Ability to interpret and present detailed account of findings and their implications	Ability identify actors, outcomes and preferences for a specific strategic or crisis situation of interest To construct and complete the set of relevant subjective decision matrices Ability to interpret and discuss results of analysis in light of strategic objectives and broader policy concerns
Diverse / Minimal	0	Individual user or team with enough familiarity with several specific techniques and / or software applications to be able to implement an existing model and answer a pre-defined research question	Ability to run basic statistical analysis in SPSS or another statistics package and link the interpretation and implications of those results by to a specific structured case study.
Diverse / Average	3	Individual user or team with sufficient experience and training with several specific techniques and /or software applications Ability to take a well-defined research question, build and run appropriate individual models and provide basic report of findings	Ability to construct and complete a subjective decision matrix and use those results to inform the design and information input for a table top war game Ability to analyze the results in light of a pre-defined research question
Diverse / Advanced	4	Individual user or team with advanced training and considerable expertise with several specific techniques and /or software applications, ability to program software as needed Ability to define appropriate research question, build and run a multi-method model to address that question Ability to interpret and present detailed synthesis of findings and their policy implications	Ability identify actors, outcomes and preferences for a specific strategic or crisis situation of interest Ability to construct and complete a subjective decision matrix and use those results to inform the design and information input for a table top war game Ability to run war game, analyze results and apply to broader strategic and policy context

GENERIC MODELING TECHNIQUES

AGENT BASED MODELS

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ❑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ❑ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

One of the most commonly used approaches to understanding the emergent behavior of human groups is the Agent-Based Model (ABM). ABMs are a type of simulation that employs a bottom up approach in which heterogeneous agents, agent characteristics, interaction rules, and the environment are explicitly modeled resulting in emergent complex social phenomena. The critical feature of emergent phenomena is that they are system-level patterns that result from the interaction of agents but that cannot easily be derived solely from knowledge of agent attributes and behaviors. The agents are simplified versions of real-life counterparts (e.g., ants, people, robots, or groups), retaining social and cognitive features relevant to the phenomena of interest. Agents are typically endowed with autonomy (they are not under the control of others), social ability (they can interact with each other), reactivity (they can perceive and respond to the environment), and proactivity (they can enact goal-directed behaviors). Agents interact in a virtual world, constrained and enabled by their spatial or network position.

METHODOLOGICAL OVERVIEW

ABMs are particularly useful for exploring contingencies, domino effects, and what-if analyses. They can be used both to *anticipate* system behavior and to *project* how the modeled system will respond to a particular event or intervention. An example of the type of question for which ABM is an appropriate analytic method is:

> *What are the possible effects on local public support for insurgent forces (a collective property of a population of agents) of increasing the ISAF footprint on the ground in Helmand Province (a change in the agents' environment)?*

Because system level features emerge from the complex interactions among agents and their environment, ABMs can produce non-intuitive results. This can be one of the most beneficial features of an ABM, as it assists the analyst in evaluating the validity of his or her assumptions and beliefs about the workings of a complex system. One should resist the temptation to "game" the model by tweaking the input parameters so that system behaviors fit intuitive expectations. A good ABM will simplify the problem to its essential components and provide a clear explanation of the logic of the model to ensure that the challenge of tracing the behavior of the system and understanding what led to the model's results is not an intractable problem.

Like other types of simulation, most ABMs include some probabilistic elements. For example, the information an agent receives, the agents with which a given agent interacts, the decision that an agent makes, or the effect of an action taken by an agent may involve "rolling the dice". Importantly, the options that are available to a given agent at a given time are typically constrained by the agent's individual characteristics and history. These probabilistic events affect each agent's trajectory within the model. Because such trajectories are probabilistic and not deterministic, a single run of an ABM represents only one possible way the system might evolve. **Proper use of ABMs requires many "runs" of the model and analysis of the trends across different runs. The specifics of a single ABM run have no independent analytic value.**

HOW IS THIS DONE?

The four main components of an ABM are

- Agents that act according to differing attitudes, social cues, norms, etc.;
- The relations between agents or agents and the environment that influence their behavior;
- The actions that agents can take and the behavioral rules that govern their actions; and
- The environment in which the agents exist and the relations between agents and selected environmental factors.

To apply agent-based modeling, the analyst builds an ABM framework by specifying the agents, the environment, and the interaction logics. It may take several iterations of design and testing to ensure that the behavior of the model is true to the real-world scenario under investigation. Because of the complex interactions that take place in an ABM, it may be difficult to identify which model features are related to observed outcomes at the system level and, thus, to debug models that are not functioning as expected. When the model has been validated, the analyst can design virtual experiments—systematically varying several variables and running the simulation for some predefined number of time steps while monitoring and measuring select system parameters. Multiple replications per experimental design cell in the experiment are conducted to capture the range of possible outcomes for any given set of initial conditions. The results are then analyzed statistically to identify trends, impact(s) of changes in the independent variables in the experiment, and overall behavior.

ABMs are widely used to evaluate the impact of interventions on groups, cities, nations, and the world; identify future possibilities due to socio-economic changes; assess the impact of removal of key leaders; and so on. Essentially, **ABMs are part of the standard toolkit for making forecasts and doing counter-factual reasoning**. Any group of three or more agents can be meaningfully assessed using this technique, regardless of the kind of agent. Agents can be individuals or groups, people or animals, or some phenomena (such as a power source) that can affect the environment.

Interpreting and Using ABM Results:

*Clearly identify the discrete purpose for which the ABM was built***.** There is really no such thing as an all-purpose ABM. Like all quantitative and computational models, ABMs are simplifications of small portions of reality. Taking the example question above, an ABM built for the purpose of exploring the effect on opposition support of increasing coalition forces in Helmand would not necessarily contain the detail on agents and rules required to also explore the effect on insurgent support of improved healthcare in Helmand.

Use ABM results to provide insight not "answers." An ABM is intended for gaining insight into the dynamics of a system comprised of many interdependent actors. They are not suitable (or even intended) for predicting the occurrence of a specific event at a specific place and time. Consequently, care must be taken in interpreting the results of ABMs as suggestive; i.e., as highlighting possible outcomes of actions in terms of expected trends, rather than providing definitive, specific "answers." Moreover, although ABMs may offer guidance about events that are more or less likely, given a particular set of initial conditions, it is critical to acknowledge the limiting assumptions that have been established by the constraints of the model design, as noted above.

Consider initial conditions. ABMs can be sensitive to initial conditions and seemingly minor alterations in agent interaction rules. Thus, sensitivity analysis, in which the input parameters are systematically varied and the model is re-run, should be used to test the robustness of the model's results.

Question the input data. The precision of an ABM is only as good as the input data. ABMs must sometimes estimate the attributes of specific agents based on statistics about the agent population. In these cases, the level of analysis for which the ABM results are valid is limited by the data of lowest resolution (i.e., the grossest or least-detailed data). For example, in an ABM that models the population of a country composed of provinces that are subdivided into districts, if some of the population data is available at the province level only, then the results of the ABM should be interpreted at the province level only, even if other data is available at the district level.

REQUIREMENTS

DATA

High (4)

In order to build and populate an ABM, the types of data required include

The set of agents and their attributes;

1. The relationships between agents;
2. Historical data regarding past behaviors of the agents that can be used to calibrate and validate the model.

Modelers often glean information to build the model from:

- Subject matter experts, particularly for identifying which agents to include in the model and providing subjective estimates of attitudes, preferences, and other attributes that are cannot be measured directly;
- Documents and reports
- Event databases (e.g., SIGACT databases)
- Link analysis diagrams and databases
- Statistical data (e.g., surveys, censuses, economic data, etc.).

SET UP TIME

Average (3)

In general, the level of analyst involvement depends on the level of specificity required. The overall length of time required for the simulations depends on the size and complexity of the group being assessed. Designing the initial hypothesis and model takes weeks to months depending on the familiarity and expertise of the person or team designing the model.

The data collection phase may take several weeks if starting from scratch and the information is either readily available or if the designer has access to subject matter experts who can provide the information required.

EXECUTION TIME

Short (0.25)

Each run of the model takes from 10s of minutes to several hours depending on the complexity of the question and the detailed output required.

SKILL SET/EXPERTISE

Diverse-advanced (4)

Working familiarity with the scientific process, to include the ability to log and recreate efforts over time. Experiment execution, re-execution, and duplication are key to supporting claims.

Working familiarity with ABM software (i.e., AnyLogic, Repast, etc.), especially the inherent assumptions made and parameterization of simulations.

Elicitation skills and data discover skills (e.g., for SMEs, knowing who to ask for help, how to ask for help, which questions to ask, how to translate answers into the model; for data sources, knowing where they may be [e.g., classified networks], who grants access to sources, how to query sources, and how to translate query results into the model). Working statistical analysis skills to analyze outputs of simulations to accurately state both numeric and non-numeric conclusions.

Thorough command of domain specific language (e.g., to ensure comprehension of SME provided information as well as communicate more effectively with the client/customer). Excellent written and oral communication skills to allow communication of results at levels of detail ranging from the executive summary/abstract to the lab-engineer in both time constrained and non-time constrained situations.

TOOLS

Specialized (2)

Agent Based Modeling requires ABM software.

COST

$$$ (13)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Going back to the 5D framework, discuss how the general questions identified in the 5D can be effectively addressed using this method.

- How effective is current US force posture for achieving policy objective?

ABMs allow the analyst to address a piece of this question; specifically, given the current US force posture and, therefore, lines of alliance and conflict with other countries, how likely is it that the nations of interest will move in a direction in support or opposition of the US policy. What would be the optimal force posture to achieve a specific policy objective?

ABMS are not optimization tools. It can be used to generate a suite of hypothetical futures that could then be assessed to identify those most in keeping with a specific policy objective. The outcomes of this model could be put into an optimization framework. To assess this question, a set of alternative force postures and the associated belief structures would be represented in the model. Then the interventions associated with those force postures would be run during a virtual experiment. The results would then be assessed to identify plausible futures of interest. It should be noted that enumerating all possible interventions of interest associated with a force posture is generally not feasible, so identifying an optimal force posture is less likely than is identifying the relative strength of alternative force postures.

- What strategy is optimal to achieve the objective?

To use ABMs to answer this question, the analyst would need to specify a set of strategies, run a virtual experiment where these alternative strategies were run, then analyze the results statistically to determine if any of the strategies outperformed the others with some degree of significance. We note that, in general, optimality is often not the goal; rather, the objective is to identify a set of strategies that meet the objective so that factors external to the model can be used to choose between them.

Agent-based simulations are, by nature, abstractions of the real world. Agents in ABMs have limited cognitive capabilities. They are not humans. Agents have no emotional capabilities and have, presently, fairly limited goal/task-oriented capabilities. The simulations' results are therefore applicable to emergent and collective behavior analysis, not specific-agent behavior analysis. Attempting to discern why a particular agent performed a particular way at a particular time is not a question ABMs are equipped to answer with medium or high confidence.

Simulations that operate in multiple dimensions can be difficult to fully grasp for humans (technically oriented humans as well as lay-people). The very attempt to capture complexities of real life human interactions within the simulations can make communicating both the design and the output of the simulation difficult, potentially degrading the confidence of clients/customers in the results communicated to them by analysts.

Interactions of inputs, as well as parameters of the simulation, are almost always non-linear. This may be the essence of "emergence"—the fact that there is no straightforward way to connect the aggregates actions of individual agents to the collective behavior of the system. Fine tuning a model, akin to changing an equalizer of a stereo system, can lead to discovery of inflection points and potentially discontinuities in outputs that are not immediately discernible or anticipated by analysts or clients/customers. As such, it is less likely that a question of "How much of X do I need to cause Y" will be easily answered using an ABM if X and Y interact with other inputs or outputs.

FURTHER RESOURCES

General ABM Resources

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academies of Science*. Vol. 99, supl. 3. Retrieved from http://www.pnas.org/content/99/suppl.3/7280.short

Macy, M. & R. Willer. (2002). From factors to actors: Computation sociology and agent-based modeling. *Annual Review of Sociology*. Vol. 28. Retrieved from http://www.jstor.org/pss/3069238

Macal, C.M. (2010). Tutorial on agent-based modeling and simulation. *Journal of Simulation.* Vol. 4. Retrieved from http://www.palgrave-journals.com/jos/journal/v4/n3/abs/jos20103a.html

Gilbert, N. (2008). Agent-based models. Series: Quantitative applications in the social sciences. California: *Sage Publications.* Retrieved from http://books.google.com/books?hl=en&lr=&id=Z3cp0ZBK9UsC&oi=fnd&pg=PR9&dq=agent+based+modeling&ots=T3FExDnUmY&sig=zYcAJpzwBy6skdK1zHbT2mNbJy4#v=onepage&q=agent%20based%20modeling&f=false.

Resources Specific to Nuclear Deterrence and ABMs

Kyungkook, K. & J.B. Compton. (2008). Testing deterrence: An agent-based modeling approach. *Claremont Graduate University, California*. Retrieved from http://handle.dtic.mil/100.2/ADA500918

SUBJECTIVE DECISION ANALYSIS

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ❑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ❑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ❑ **Time-series** : multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** fewer than three observations or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ☑ **Qualitative**
- ❑ **Quantitative and qualitative**

INTRODUCTION

A major contention of decision analysis is that the subjective processes by which decisions are made affect the choices that individuals make and, thus, the outcomes that follow from those choices. In other words, we cannot explain or predict behaviors without understanding how individual and group perceptions, values, and preferences are transformed into decisions. While it is virtually impossible to predict the specific action that results from an individual's perceptions at any given moment, there are factors that can help us to understand behavioral responses to certain types of decision processes (Astorino-Courtois 1998). The subjective decision analysis technique discussed below provides a way to "reconstruct" an opponent's decision problem from his own perspective so that the effects of his perceptions, particular interests, and understanding of the decision itself can be systematically examined. There are other methods used in subjective decision analysis, such as process tracing and experimentation; however, both of these require significantly more resources. Additionally, subjective decision analysis can be done prospectively and at a distance, which increases its utility for analysts and planners.

METHODOLOGICAL OVERVIEW

The decision analysis process formalizes what an actor perceives his options to be in a particular decision setting. Most types of non-myopic decision analysis (e.g., rational actor and others discussed below) highlight 1) the primary interests an actor seeks to maximize; 2) who the actor perceives as its opponent, or the "other player or players" involved in a decision problem; and 3) the actor's assessment of the potential actions or responses to his decision that the opponent will consider. In subjective decision analysis, these factors are included but rather than, for example, assuming non-emotive or value-neutral interest maximization, the actor's own perceptions, values, and emotions, and the ways in which he receives and processes information about these are included.

These aspects of an actor's decision calculus can be formalized into a decision matrix that enables the analyst to use various choice rules or decision heuristics; for example, choosing the first option that appears to a decision maker to be "good enough", to explore what changes in environment or perceptions might drive an actor toward or away from a particular choice option (e.g., to lay an IED or not). Depending on the fidelity of the data available, subjective decision matrixes can be built to model decision makers at the level of a particular individual, a group, or larger leadership. The unit of analysis should be selected before beginning construction of the full matrix. In some cases, divisions of power within a group (for example military versus political leadership within a country) require that separate matrixes be constructed.

ASSUMPTIONS

All modeling approaches are based on a set of assumptions that simplify the process under examination. Subjective decision analysis makes the following assumptions regarding decision makers:

- Decision makers have distinct interests that can be treated as discrete (or that otherwise aggregate into a single interest).
- Decision makers can identify, at least within their own realm, their available options in a given decision context.
- Decision makers can and have identified that a decision needs to be made.
- Decision makers act *as if* they were using decision calculus matrices.

HOW IS THIS DONE?

A subjective decision analysis approach seeks to build an understanding of how an actor's interests and perceived options, as well as those attributed to other actors, affect its decision-making process and behavior. Subjective decision analysis is a systematic, qualitative approach based on three steps.

1. Construction of subjective decision calculus matrices for all actor(s) and decisions to be studied.
2. Identification of the possible manipulations that can make an option appear more or less advantageous to a specific actor.

3. Creation of an inter-subjective decision model using decision matrices linked according to the decisions and issues.

Constructing a Subjective Decision Matrix

This step formalizes the decision calculus of an actor by identifying interests and perceived outcomes. The process begins with the collection of source material on the issue to be examined. The analyst uses this material to identify:

- Principal decision makers
- Policy alternatives considered by the decision makers
- Policy alternatives attributed by the decision makers to other actors involved in the event
- The interests or value dimensions considered by the decision maker
- The utility and probability weighting of outcomes (to the extent that they are made)
- The ranking of perceived possible outcomes

This process also serves to systematize the analysis across different actors and settings, which enables the analyst to directly compare across actors and decisions. Furthermore, once an analyst has completed the data generation process for one case, the same skills can be applied to other cases or actors.

The data collected is then used to construct search-evaluation (SE) matrices for each decision maker. As shown below, SE matrices are a graphical representation of a multidimensional problem space. The rows list the decision outcomes identified by the decision maker, judged across the dimensions (columns) identified as relevant to that decision maker in that decision context (Maoz 1990). It is important to note that the SE matrices are a reflection of the decision-maker's *perception* of the problem space, not an objective mapping of the problem space or the analyst's own perception of the problem space. They may, therefore, not contain all the possible outcomes of a specific problem or all the dimensions the analyst, prior decisions, theory or SME opinion may consider relevant[2].

Once the matrix is built, the outcomes in the matrix must be evaluated and then ordinally ranked according to the degree to which they satisfy the decision-maker's preferences on each of the identified dimensions. These single-dimension preferences can then be aggregated across the set of dimensions for each outcome to produce a multidimensional preference ordering for the entire choice set.

[2] For further discussion and examples of SE matrix construction and analysis see Maoz 1990.

Actor A Choice Options (US opponent)	Options attributed to Actor B (US)	Actor A Interest 1: National security	Actor A Interest 2: Domestic public opinion	Actor A Interest 3: National security	Actor A Interest 4: International reputation	Rank
Proliferate	Ignore					
	Sanction					
	Military strike					
Do not proliferate	Reward					
	Ignore					

Figure 1: Example of a Search-Evaluation Matrix

Manipulating Preferences

This step provides possible options keyed to an actor's own interests and perceived incentive structure. Once the SE matrix has been created, it becomes easier to identify the most likely means by which an opponent's preference ordering might be changed. Each option an actor perceives generates multiple outcomes, depending on the perceived predicted response of the other player. For example, in the SE matrix above, Actor 1's choice to proliferate generates three possible outcomes; proliferation with no response by the US, proliferation followed by sanctions, and proliferation followed by a military strike. Not proliferating generates two outcomes; non-proliferation rewarded by the US and non-proliferation ignored by the US. The ranking order reflects how well each of these outcomes serves Actor A's interests. Changing your opponent's perception of the response you are most likely to take, given a specific action, is one way in which the utility of a choice option, and thus the final choice, can be manipulated. Examination of the relationship between an actor's interests and specific choice options can also be informative. For example, a choice option is ranked high because it provides the greatest utility on the most important issue for an actor. In this example, Actor A's prioritization of national security may drive the high ranking of proliferation. If this is the case, a US offer of protection in exchange for a commitment not to proliferate may enable Actor A to achieve their national security goals without proliferating. On the other hand, if proliferation is driven by domestic political considerations (national pride, need to demonstrate strength to domestic public), it will be harder for the US to manipulate A's preferences such that non-proliferations ranks above proliferation.

Creating an Inter-subjective Decision Model

This step can provide insight into a regional or multi-actor decision setting. It can also uncover misperceptions, volatilities, and possible strategic advantages that result from such misperceptions. Once the analyst has constructed an SE matrix for each actor in the problem space, the matrices can be compared in order to identify inconsistencies between individual decision-makers' perceptions. This can provide information regarding potential courses of action that may lead to unintended consequences. In post-hoc analysis of a policy choice or crisis situation, this type of analysis can be a useful forensic tool in determining exactly why the outcome of a decision deviated from expectations.

Limitations

The most significant limitation of this approach is that decision matrices are not dynamic. Although decision analysis is best suited to strategic level decisions, for which the underlying incentives change more slowly than do those for tactical decisions, the static nature of the technique does not allow for easy updating. It is possible to link decision matrices, but this approach quickly becomes unwieldy.

REQUIREMENTS

DATA

Average (3)

It requires a substantial amount of qualitative primary and secondary source information to generate decision matrices. Information on the event to be analyzed, the relevant decision makers and their perception of the event and other relevant actors is necessary.

SET UP TIME

Average (3)

The majority of time needed to complete a cognitive decision analysis is involved in the collection of information. As an SE matrix is composed of information regarding a specific individual or group's perceptions of a problem space, the information needed to complete the matrix can be difficult and time consuming to collect.

EXECUTION TIME

Short (0.25)

Once the SE matrix is completed analysis and interpretation is relatively fast.

SKILL SET/EXPERTISE

Specialized-advanced (2)

As with any analytic technique, the quality of a cognitive decision analysis will be improved if it is undertaken by an individual or group that has some prior experience with the technique. In this case a familiarity with basic decision theory would also be beneficial, particularly in the design and analysis stages. Proficiency in any relevant foreign languages is very beneficial at data collection stage.

TOOLS

Minimal (0)

There is no requirement for any specific technology or software to complete a cognitive decision analysis. However, tools such as the Decision Analysis Tool (DAT) designed by NSI can make analysis of results much easier and complete.

COST
$$ (8.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Decision analysis focuses on understanding the resolution of a problem by examining the process by which actors made decisions related to that problem space. Subjective decision analysis provides a means of understanding the effects of the decision makers' own preferences and perceptions on the choice process and outcome.

This makes subjective decision analysis an excellent technique for addressing questions regarding:

- Behavioral responses to certain types of decision processes and conditions
- What drives a specific actor's choices and behavior
- What factors are central in moving an actor toward or away from a specific decision
- How divergent perceptions of a problem space among actors can lead to misperceptions and unintended and unexpected consequences.

In the context of nuclear policy, subjective decision analysis could be applied to question such as:

- How is Russia likely to respond to a US decision to pursue ABM capabilities?
- What US policy is most likely to stop North Korea from continuing to pursue nuclear weapons capability?
- How can we best assure Japan?
- If Iran continues to pursue nuclear weapons capabilities, what US actions are most likely to assure regional actors and prevent further proliferation?
- Will the US be able to deter China from invading Taiwan if it reduces its nuclear arsenal by 20%?

FURTHER RESOURCES

Astorino-Courtois, A. (1995) The Cognitive Structure of Decision Making and the Course of Arab-Israeli Relations 1970-1978, *Journal of Conflict Resolution*, 39: 419-438.

George, A. (1980) *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice*. Boulder CO: Westview Press .

Keeney, R. and H. Raiffa (1993) *Decisions with Multiple Trade-offs: Preferences and Value Trade-offs.* New York, NY: Cambridge University Press.

Maoz, Z. (1986) Multiple Paths to Choice: An Approach for the Analysis of Foreign Policy Analysis. In *Different Text Analysis Procedures for the Study of Decision Making*, edited by I.N. Gallhofer, W.E. Saris and M. Melman, pp. 69-96. Amsterdam: Sociometric Research Foundation.

Maoz, Z. (1990) *National Choices and International Processes* Cambridge: Cambridge University Press.

Payne, J.W. (1976) Task Complexity and Contingent Process in Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance* 22: 17-44 .

Simon, H. A. (1956) Rational choice and the structure of the environment. *Psychological Review*, Vol. 63 No. 2, 129-138.

COMPLEXITY-BASED APPROACH

Identify the levels, observations, temporal domains, and forms for which this method is most suitable.

Levels

- ☑ **Global**
- ☑ **Regional/ multi-actor grouping**
- ☑ **Nation-state or non-state actor**
- ❑ **Sub-national/organization group** (e.g., Pakistani military)
- ❑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Large-N**
- ❑ **Small-N**

Temporal domains

- ☑ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Forms

- ☑ **Quantitative**
- ☑ **Qualitative**
- ❑ **Quantitative and qualitative**

INTRODUCTION

The term "complexity" appears in strategic guidance, policy statements, and doctrine with increasing frequency. Described in 1973 by Rittel and Weber as "wicked problems", complex problems are ones that defy simple explanations. They are difficult for us to comprehend because they contain many interconnected elements and subsystems reacting to each other simultaneously, making the prediction of specific outcomes practically impossible. When one hears anyone talking about "unanticipated second- and third-order effects" or "unintended consequences", they are most likely referring to an encounter with a complex system.

There is no such thing as a "complexity method" of analysis, rather complexity-based synthesis offers a frame of reference that reflects the way the world really works. New insights from complexity science can help planners understand both the usefulness and limitations of various analytical methods when facing a dynamic world, one that can never be perfectly understood nor predicted.

METHODOLOGICAL OVERVIEW

The basic theoretical underpinning in any quantitative analysis is the assumption that the world can be described by reductionist approximations in linear terms. In other words, it means that "… we can get a value for the whole by adding up the values of its parts. More carefully, a function is linear if the value of the function, for any set of values assigned to its arguments, is simply a weighted sum of those values" (Holland, 1995). Linear sciences have been invaluable to most of the technological innovations that we often take for granted today, but in the last thirty years, it has become increasingly apparent that they are not sufficient to accurately depict the level of complexity that characterizes "wicked problems." As the tools provided by the linear sciences improved, scientists' abilities to look at the parts of the world separately, those same tools, especially modern computers, started to give them new insights into the interactions of the parts as well.

But as scientists from different fields traded notes on what they were seeing in their own fields of study, it became increasingly apparent that many of the same nonlinear principles were at work in all of the other fields, ones that could not be adequately described with reductionist concepts from traditional analysis. Complexity was originally explored by the Santa Fe Institute, whose founding in 1984 and subsequent development is described in M. Mitchell Waldrop's 1992 book *Complexity: The Emerging Science at the Edge of Chaos.* The Santa Fe's interdisciplinary effort to understand discipline spanning phenomena, not easily described by linear sciences, has developed into a body of scientific principles known collectively as complexity theory.

WHAT IS COMPLEXITY?

While there is no single accepted definition of complexity, all descriptions of complex systems describe how, at the macro level, the collective properties of interconnected systems usually cannot be predicted nor understood merely by recombining understandings of the individual parts. As described by Yaneer Bar -Yam, the President of the New England Complex Studies Institute (NECSI),

> "Complex Systems" is a new approach to science, which studies how relationships between parts give rise to the collective behaviors of a system and how the system forms relationships with its environment. Social systems arise (in part) out of relationships between people, the brain's behaviors result from relationships between neurons, molecules are formed out of relationships between atoms, and weather patterns are formed because of relationships between air flows. Social systems, the human brain, molecules, and weather patterns are all examples of complex systems. Studying complex systems cuts across all disciplines of science, as well as engineering, management, and medicine. It is also relevant to the humanities; art, history, and literature. It focuses on certain questions about relationships and how they make collections of parts into wholes. These questions are relevant to all systems that we care about (Bar-Yam, 2004).

There is also no established method to measure the degree of complexity within a defined system, although there are ways to characterize relative complexity between various systems. As described by Dietrich Dorner (1996) in *The Logic of Failure*,

> *Complexity* is the label we give to the existence of many interdependent variables in a given system. The more variables and the greater their independence, the greater the system's complexity…The links between the variables oblige us to attend to a great many features simultaneously, and that, concomitantly, makes it impossible for us to undertake only one action in a complex system.

In real life, complex physical and informational systems are constantly reconnecting and recombining in new and novel ways, making them even more difficult to model or approximate. For example, war is sometimes compared to a chess game between opponents. The complexity informed perspective proposes that conflict would be better described as several games being played simultaneously - imagine western chess, Chinese Go, Indian Parcheesi, etc. being played on the same multidimensional game board, with both the primary opponents and the audience moving pieces. Some players have more pieces than the other players, and the relative values and move rules of the pieces are constantly changing depending on who is most actively attacking or defending. There is no sequential play, and multiple pieces are moved during the same turn. While all of this is going on, the players are constantly communicating with each other through a series of implicit and explicit messages and alternately cooperating or competing with each other. While some players announce some of their moves, they hide others and alternately bluff and tell the truth about their intentions.

As complex and chaotic as this imaginary contest seems, in truth, it is a gross oversimplification of what really happens every day in the global security environment. Given the significant amounts of computing power needed to defeat a skilled human opponent in the formally defined two-dimensional rules of chess, this thought exercise should serve as a cautionary tale as we apply linear-based tools of analysis to highly complex matters, such as nuclear deterrence and assurance.

PREDICTING OUTCOMES

One of the keys to understanding the nature of complexity is the difficulty in predicting specific outcomes of a single action in a complex system. The more highly connected the variables of a system, and the more sensitive various parts of the system are to perturbations in other parts, the more difficult it is to anticipate outcomes from a single act, let alone multiple acts within the system. In certain combinations of conditions, a single or small number of variables has a disproportionately large influence on the way the system reacts. How much even slight perturbations can drive the system from conditions of stability, an inherent potential for rapid-phase transitions in some complex systems, is commonly referred to as the "butterfly effect". Complex systems make it difficult to predict the outcomes of operations, as single actions often create multiple cascading outcomes that we did not intend or anticipate. Thus, complexity is not only a description of the interconnectedness of the parts of a system, but also our cognitive ability to account for it; it describes a relationship between the observer and the system being observed.

THE NATURE OF SOCIAL PHENOMENON

Complexity-based synthesis recognizes that there are both useful quantifiable and unquantifiable aspects of nearly any social problem and contends that neither purely quantitative nor qualitative approaches are sufficient unto themselves. We frequently use quantitative measures to describe both concrete characteristics of individuals (e.g. income, education) and more qualitative characteristics (e.g. attractiveness). While many of these relationships may indeed be correlated with emergent social properties, such as success and popularity, none of these measures can ultimately predict emergent psychological states that may make certain behaviors more likely. For example; "Is this person happy?", or "Is this couple in love?" are questions are entirely dependent on the individual context, and the subjects themselves may not even be able to answer the question. This points to what might be the fundamental limit of analytical modeling of human behaviors. Even if we can approximate human decision-making by creating quantitative models capable of approximating stochastic processes and randomness, those models would likely retain the same amount of unpredictability and irrationality as humans themselves.

INTERACTION BETWEEN SYSTEM PARTS AND THE WHOLE

Complexity-based thinking cautions that sensitivities and properties can only be understood in the context of the whole of the system working together. When an analytic technique holds some variables constant in order to examine the relationships between individual parts, our understanding of the whole may be compromised. For example; think of a tornado. One could study the properties of millions of different moving objects in one square mile area individually, naming color, size, hardness, cellular structure, and relative velocity to one another, but in terms of understanding the implications of phenomenon itself, it is far more useful to us to look at the properties of system in aggregate and say "There is a tornado heading straight for us, time to get in the basement". These are **emergent properties** of the system, or properties that can only be understood in the context of aggregation and flows over time (Holland, 1998). The concept of emergence can be applied to physical systems like a tornado or also to individual and group attitudes and behaviors that emerge from the interaction of people sharing ideas. This makes the concept of emergence useful in considering issues like deterrence, assurance, and stability.

STABILITY THROUGH THE LENS OF COMPLEXITY

Most generally, **stability** describes how much of a shock a system can absorb without being sent into a different state of equilibrium. In a social sense, stability - like complexity - describes a relationship between the observer and the system being observed. It is a term we use to describe our comfort level with the rate of change in the world as we perceive it, which is partially determined by the degree to which we can anticipate, shape, cope with, and accept change. While our perceptions of stability are indeed an emergent property of individual and collective social consciousness, stability in this sense is not a physical property of the social system that can be objectively sensed or measured.

Stability indicates the presence of **attractors**, physical or cognitive components of the system that guide the direction of movement in that system. The attractors that encourage stability come in many forms; they can be physical features that limit or drive patterns of human activity, such as natural barriers, like mountains, rivers, and oceans. They can be conceptual constructs, like the laws and norms that establish and legitimize political and economic reward and punishment systems. They can also be informal codes and norms which are not recorded in the official social hierarchy or written codes, yet still drive individual and group behaviors (i.e. "honor among thieves"). These attractors lead to somewhat predictable patterns of interaction in the physical world, forming the compensating positive and negative feedback loops that produce relative stability over a given period of time. Stability does not describe statis, as evolution and growth depend on some degree of change in the system. When society works well, change proceeds with sufficient hope, opportunity, and basic necessities to satisfy the needs of the majority of the group.

The outward appearance of stability can be deceiving, however, as it can hide changes that are gradually making the system more vulnerable to destabilization from small disruptions. Not everybody benefits from stability; those who are unhappy with the status quo want change and will deliberately seek to create instability to force change, either when it is the most expedient course of action, the only course of action, or both. Conflict at lower levels does not necessarily mean there is not stability at higher levels. Proxy wars, for example, show that ideologically-opposed nations can compete through secondary nations or groups, while preserving higher level stability. In such scenarios, stability derives from either stalemate or the mutual desire to avoid a negative outcome.

ADVANTAGES OF COMPLEXITY-BASED APPROACHES

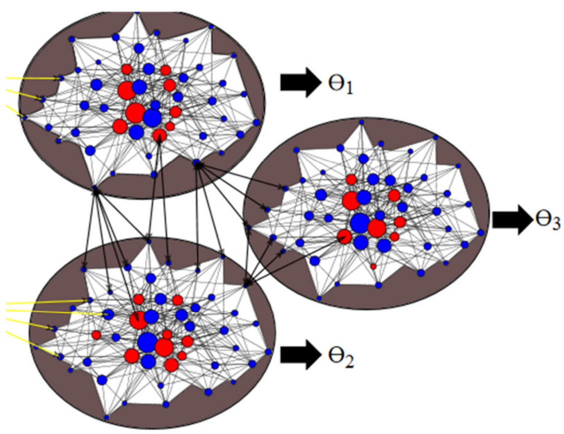
Given the significant challenges to understanding and prediction, like emergent social properties and multiple causes for individual actions, how should the practical strategist and planner proceed? While complexity-based thinking helps us to understand the limits of our understanding, it also suggests ways that we can take advantage of the same forces of emergence and complexity within our planning groups, setting conditions that encourage the formation of an emergent collective intelligence that exceeds the individual intelligences of its members.

LEVERAGING OUR EXISTING COGNITIVE CAPABILITIES

When conducting analysis for the purpose of strategic planning, it is important to state up front that in a fundamentally unpredictable world, one made so by its adaptive nature, the purpose of analysis is not to attempt to predict specific future outcomes. Rather, it is to help decision-makers better understand how the world works, or more specifically, how it adapts so that they can better anticipate outcomes, understand the limits of control, and use what control they have to shape future events in favorable ways. Rational strategists only seek the achievement of specific ends if they perceive that those ends will serve to promote a continuing advantage in the patterns of societal adaptation (Dolman, 2005). This requires structural understanding of how the moving parts of society are interconnected and also a contextual understanding of how these movements will be processed and interpreted by other adaptive agents. Humans do this by literally constructing mental models of the world in our own brains, models formed and defined by changing associations between many billions of neurons, each with thousands of parallel connections that can be traveled in countless ways, giving pattern recognition and association capabilities that far exceed the capabilities of any manmade computer (Morgan, 2006). Thus, the real purpose of analysis and planning is quite literally to inform and improve the models of the dynamic world in the heads of key decision-makers so that they better understand the possible implications and risks from their decisions (Gardener, 2006). The better their mental models, the better they can anticipate and recognize key patterns in the real world, helping them to choose courses of action that are likely to create more favorable outcomes than negative ones.

DESCRIBING COLLECTIVE MENTAL MODELS

Just as parts of our individual mental models are stored in different parts of our brains, parts of our collective mental models reside in the individuals that make up that group. Each member of a group has core elements of that group's collective mental model in their own model. We describe these common understandings as culture, norms, mores, and laws, but each member of the group also has individual knowledge from their unique experiences that can benefit the larger group model, if properly harnessed. If one uses the metaphor of shared mental models, the following graphic model can describe how the interaction of three individuals with different individual models (represented by the configurations of blue and red dots) moving forward through time influence each other to develop a similar collective mental model (the shape of the white outline that is approximately common to all three) (Duda & Levinson, 2011).



Alternately, the diagrams above could be interpreted to show how the interaction of individuals in three different groups (in the green circles) inform the formation of similar collective mental models, with various individuals having different degrees of pull on the direction and shape of the collective group model. The bigger red and blue circles could be the people with the most social clout, either by force of personality or by position within the social system. The multiple uses of a network depiction like this are evidence of one of the additional benefits of using complexity science to understand complex problems; you can translate metaphors across disciplines and levels of scale to form better mental models of what is happening in your own head. What does this diagram really depict? In actuality, it shows the linkages of individual groups of neurons and how audio and visual cues are combined to form an individual's perception of outside events. The translatability of the model comes from the common properties inherent in any type of network. As Stephen Johnson points out, ideas are actually networks of neurons with the semi-permanent structures providing memory, and changeable elements updating the idea as new perceptions are sensed, classified, and interpreted (Johnson, 2010). If one used the "people as linked dots" interpretation of this basic model of networks, you could use this diagram to explain how all three of Graham Allison's political decision-making models (Rational Actor, Organizational Behavior, and Governmental Politics Models) work at once (Allison & Zelikow, 1999).

THE WISDOM OF CROWDS

Recent studies have shown that when information from many people is brought to bear on questions, the "Wisdom of Crowds" can smooth individual errors in judgment and apply multiple perspectives, in effect, providing the missing pieces of the collective mental model (Surowiecki, 2004). This facilitates more holistic understandings of complex problems and sometimes produces very precise estimates of the more quantifiable elements of the system when the question posed is specific enough. For example, averaged estimates of estimated weights and the numbers of objects in jars are uncannily accurate when the group polled is large enough. In a sense, both individual brains and the collective brains of groups act like a biological computer, offering resilience by storing critical information in many places, which allow societies to survive systemic shocks, such as pandemics, financial crises, and physical destruction from disasters and wars.

COLLECTIVE MEMORY

Just as in the human brains, in societies, ideas are stored in a combination of semi-permanent and highly-fluid structures. The most permanent non-physical structures that span multiple contexts, we call stories, narratives, and mores. The more specifically applied and less permanent manifestations are laws and customs, and the physical manifestations of ideas become embedded in bureaucracy, including rewards and punishment systems. It is this combination of physical and conceptual architectures that determine the resilience of some ideas over others; ideas embedded in bureaucratic structure and processes gain a type of inertia that resist changes as a necessary function, as the purpose of bureaucracy is to provide reliable and predictable processes. Understanding how collective ideas and physical structures relate improves our ability to anticipate possible outcomes; complexity is not randomness or chaos and depends on some parts of the system remaining stable, even when on the "edge of chaos." In terms of complexity, these established physical characteristics, ideas, and social structures serve as attractor, and help to set the limits of social adaptation, reinforcing certain behaviors while discouraging others.

If one understands what the physical and conceptual attractors are in social systems and how they work with other attractors dynamically, a capable strategist can seek to control, change, destroy, or create physical and conceptual attractors in order to create advantage. At the very least, when insufficient ways and means are available to positively shape events, the understanding of the way attractors work can help the strategist to anticipate future patterns of social adaptation, and seek ways to mitigate other's ability to impose unfavorable conditions upon them.

HOW CAN COMPLEXITY-BASED SYNTHESIS BE APPLIED?

Using complexity-based thinking to better understand the strengths and limitations of other analytic methods can help one gain insights from them while minimizing their limitations. By understanding the nature of the assumptions inherent in a particular analytical method, one can deduce which factors or combinations of factors might be most relevant in specific situations. Insights from both quantitative and qualitative analysis must ultimately be synthesized in order to provide decision-makers useful insights. These insights must be effectively communicated in a manner that is accessible to the intended audience, namely the senior leaders and policy makers that will use this information to inform their mental models for the purposes of interpreting the world and making decisions.

No matter what degree of fidelity we can bring to analysis and synthesis, and regardless of how powerful our "bottom up" collective insights are, the structure of bureaucracy still leaves large power in the hands of those in key positions within the social structure. It is ultimately the internal mental models of a relatively small number of key players that will have the greatest impact on determining the policies and decisions that affect the lives of billions of people, given the leader-centered structures inherent in most military and governmental bureaucracies. This usually presents the advantages of consistency and predictability in policy, but also the risk that the choices these key players make with imperfect understandings of complex systems may have far reaching unintended negative consequences. Our goal to hedge against this is to build on key leaders' existing understandings of the world and enhance them with the collective wisdom of our group effort, condensed into the forms of decision briefing presentations and recommendations.

As it is suggested by the assertion that "Planning is more important than the plan", the most important part of this process is not relaying the recommendations themselves, but rather conveying the logic that led to them. It is the insights into the drivers of complex system dynamics, not specific predictions of outcomes, which allow leaders to anticipate change and adapt their policies to an uncertain and unfolding future. Unfortunately, due to the competing demands on the limited time available with these leaders, the insights must be distilled and are usually presented in two-dimensional PowerPoint slides and bullet statements that do not make these dynamic insights intuitive. This speaks to the importance of finding new ways to show dynamic, complex relationships in a manner that uses motion, color, and visual aggregation of data to help them relate the insights of the study to their own previous experience in a manner that conveys information very quickly and intuitively.

Complexity-based synthesis is consistent with current military planning processes, like the Joint Operations Planning Process, and can also be implemented within the Design process recently introduced by the US Army School of Advanced Military Studies. Both propose a conceptual framework and team makeup that will make those more effective and allow them to tie in the results of multiple analytical tools.

Analysis of general situations can be useful to understanding how pieces of the overall system are related, but general studies do not sufficiently account for the specific and unique circumstances that often decide when stability yields to instability or decisions are triggered in individuals or groups. For example, running a fictional candidate with notional personality traits against a real sitting president in a straw poll may tell you something about people's preferences, but once you place an actual candidate in the poll, the results may change dramatically, even if the actual candidate has the properties described by the fictional candidate. The only way to explore these complex sensitivities is to run a series of iterative wargames that test various scenarios and their underlying assumptions under specific conditions that would likely yield these types of sensitive results. This also focuses the collective mental models of the planning group on a narrower set of real world possibilities, increasing the chance that you will detect actual opportunities or deficiencies within that specific case.

- Define specific geographic and social contexts for inquiry involving the key nations and groups who have the most influence over issues of nuclear deterrence, assurance, and non-proliferation. Run each scenario step by step, seeking group opinions on the reactions of each actor after each move. Specific questions will allow for much better results when applying the "Wisdom of Crowds" to gain insight and assure that the assumptions of various subject matter experts are addressing the same problem sets and underlying assumptions. Seek group insights as to which actions or combinations of the actors reacting to one another are the most likely to contribute to a system change – in other words, when do the attractors that usually maintain stability (i.e. economic interests, internal affairs) break down, allowing new attractors (i.e. desire to maintain national honor, external considerations, etc.) to set the parameters that dominate the new status quo?
- Run the same scenario with different sets of assumptions (i.e. different schools of deterrence thought, different enemy force configurations, different political and economic situations, etc.) to see how different combinations and dispositions of military force structure provide capability to influence outcomes. Then run that same set of assumptions through multiple sub-iterations, this time with different combinations of nuclear, conventional, space, and cyber force structures available, including anticipated capabilities which may not be available but are projected to be possible in the future (i.e. directed energy weapons, extended range cruise missiles, etc.). Use these sub-iterations to identify which combinations of force composition and disposition provide policymakers with the greatest and least ability to influence outcomes, given underlying assumptions. The insights from multi-player game theory may be particularly useful for helping to sort out which combinations of potential outcomes
- Create tensions between competing actors and subgroups
- Create internal tensions within individual actors and subgroups
- Produce incentives to risk external tensions in order to relieve internal ones and vice versa
    - Which combinations of outcomes maximize our freedom of action while minimizing freedom of action of potential challengers (Beaufre, 1967)
    - After running these multiple iterations, look at the aggregated results and ask the group the following questions
    - In which types of scenarios did we see failures of stability and what caused them? Were those causes common across multiple scenarios and iterations?
    - Which individuals or groups are likely to have the greatest impact in various scenarios explored and are there certain scenarios where some of these are likely to be extremely important or almost irrelevant?

  - What roles did geography, time, and distance play in our ability to respond to certain scenarios, and are there certain combinations of forces and conditions that leave us with no options because of the geostrategic situation?
  - Which scenarios left us with few or no acceptable options? How likely are we to encounter these scenarios in the real world? How could we have prevented them by acting differently with the forces we have? Would different forces or configurations of forces have made a difference?
  - What factors outside of our control (i.e. government messaging, third party support or opposition, etc.) had the most dramatic impact on our freedom of action, and how can we manage them indirectly if we cannot control them directly?
  - Which capabilities do we not currently have, but will need, to preserve freedom of action in the future in the most likely and most dangerous scenarios?

- After these questions are considered, the planning team leaders must look at all of the combinations of requirements and capabilities, make subjective decisions about how they should be prioritized, and prepare a summary of the deliberations, the recommendations of priorities, and the assumptions that support both the assessments and recommendations.

- Present the results of the analysis in a format that cues decision-makers into what key sensitivities and assumptions undergird the analysis and synthesis. Make the assumptions behind any analytic models transparent and show the results of various iterations with different sets of assumptions. Whenever possible, use visually accessible tools that show the key dynamics in play over time and can show aggregated data in an easily accessible form. The visualization should show the interactions of the key players represented graphically in respect to specific situations and have visual cues that help decision-makers rapidly identify transition points by visual inspection, the same principle behind a dashboard warning light, but in this case showing things like red line crossings, deterrence failures, etc. which can then prompt a more in-depth examination of what combinations of factors caused the transition. A practical example of such a decision support tool in other fields of study would be an animated weather map displayed on CNN's "Magic Wall." The dynamic use of colors and motion can tell a decision-maker in seconds what the overall trends are, and the geographic context of the map instantly gives the context needed to decide whether or not weather is likely to be a factor in the areas being considered. When more information is made, a few mouse clicks can quickly access the specific data that informed the transitions noted in the macro display.

REQUIREMENTS

Effective strategy cannot be limited to military actions alone; a comprehensive approach requires skills from the full range of academic and scientific disciplines. Holistic understanding demands that planning team members use multiple viewpoints, tools, and techniques to create assessments that offer realistic projections of opportunities and risk.

DATA

Very High (5)

Complexity based synthesis can consider any relevant data that describes past and current states of the operational environment. To conduct this inquiry, planners will need historical records, detailed intelligence products, collaboration and decision support tools, specific blue and red force capabilities and requirements data. The data requirements for any type of analysis that takes this approach are considerable. Complex systems approaches seek to overcome the limitations of linear analysis techniques, by considering, simultaneously, many of the variables that define a system. The ability to achieve this demands very large quantitative data sets (either simulation or real world data), both in terms of the number of observations and the range of variables measured for each observation.

SET UP TIME

Long (4)

The data collection process is the most time consuming element of this approach. Before data collection can be undertaken, the specific boundaries of the research question must be mapped, as it is these that should determine the scope of the data collection effort. The time required for design is determined by the depth of historical understanding of the planning team members, and the extent of the problem under consideration.

EXECUTION TIME

Long (1)

CAS approaches rely on running multiple iterations in order to explore numerous possible futures and combinations of conditions within specific futures. Available computing power and the size of the data set to be analyzed will determine how long the analysis phase takes. Once the initial design and data set-up are complete, however, the analysis and interpretation of results is a much quicker process, no matter what the scale of the project. Once analysis is complete, the presentation of results in a format suitable for decision makers to understand and absorb must still be completed. Once again, project scope and computing resources will largely determine the time required for this phase.

SKILL SET/EXPERTISE

Diverse-advanced (4)

At minimum, each team should have the following skill sets:

- Subject matter expertise in region and topic of interest
- Data location and collection expertise
- Project management and planning skills
- Knowledge of statistical and other data analysis techniques
- Expertise in presentation of complex information and data analysis results

No single person has sufficient expertise to perform an adequate assessment of complex systems. Planning for action in complex environments requires personnel with diverse sets of skills – or put another way, different conceptual pieces of the emerging collective mental model - help to ensure that the group can work quickly and effectively to describe the environment, define problems, and offer solutions that are both technically and conceptually feasible.

TOOLS

Specialized (2)

Complexity based synthesis considers the insights of all relevant tools and data sources available to build the most comprehensive understanding of the most significant physical and cognitive aspects of the operational environment in various possible combinations of conditions. Most of the analytic tools and models available can be put to use within a CAS research design as long as the tool or model does not oversimplify key relationships in the system.

COST

$$$ (16)

COMPLEXITY AND THE 5D FRAMEWORK

DETERRENCE AND ASSURANCE

Deterrence and assurance can both be seen as subsets of stability and in many ways are opposite sides of the same coin; both are about offering incentives to refrain from doing something, albeit from opposite ends of the spectrum of violence. Both are designed to keep the overall patterns of adaptation trending in ways that favor the deterrer/assurer, and assume that there is at least one challenger (the deterred party) which is not in favor of the current patterns of adaptation (does not benefit from the status quo).

Deterrence and assurance do not describe properties of a social system that can be measured objectively, nor is it technically correct to call any particular asset a deterrent. Like stability, deterrence and assurance really describe a feedback flow between various actors, creating emergent states of mind and action among individuals and groups. Communication of the physical capability to create nuclear effects is one component of both, and the communication of intent to use them in response to aggression is the second. Thus, the effectiveness of either deterrence or assurance depends on our ability to influence the collective mental models of key players and groups in both challenger and ally states. As Keith Payne recently observed, the determination of adequacy in either capability or intent is dependent on the human audiences that interpret both; quantifiable measurement of either deterrence or assurance is not possible, nor is there a specific number of nuclear weapons that can guarantee either (Payne, 2011). Deterrence and assurance, like stability and defeat, reside in the minds of beholders who are viewing the same events and facts through very different cognitive lenses and interpreting them with different mental models. Additionally, different groups within the same larger group can interpret the same actions differently, meaning that abstracting nuclear nation states as rational actors may not sufficiently account for all of the subset groups who have the ability to affect deterrence failure; deterrence can fail at one level in a bureaucracy, but be preserved at other levels.

Analyzing deterrence and assurance by comparing only the relative nuclear capabilities of two potential nuclear antagonists is problematic at best, and dangerous at worst. It is not necessarily true that nuclear weapons are the only things that influence general deterrence, assurance, and proliferation; neither is it correct to say that our ability to deter or assure another nation can be described without considering the total international system in play, or even internal disruptions that might provoke deterrence failures. The effectiveness of deterrence and assurance is based on the sum total of the relationships of all of the countries interacting with each other and certain combinations of players being involved in different ways. Failing to explore the unique sensitivities that may emerge with multiple players involved in specific geographic contexts may cause planners to overlook significant attractors that influence stability, assurance, and deterrence.

Defeat

Defeat is the most problematic national policy goal for nuclear weapons that national policy dictates and we must consider. Often used as a bumper sticker to avoid contentious debate or the appearance of uncertainty, terms like defeat and victory without subsequent qualification define an interpretation of events, not the ends to be achieved. Both terms constitute an artificial psychological boundary line drawn at a certain point in a pattern of societal adaptation, one that continues after defeat or victory has been declared and does not guarantee that the advantage that existed where the line was drawn will continue in the future. Defeat may be defined as "provide enough punishment to get the other side to quit", but this is inherently unquantifiable, making defeat an almost impossible policy goal to either define or measure; perhaps this built in "wiggle room" explains the ageless popularity of the term, even if it presents a conundrum for the serious strategist.

In accordance with the 5D Framework, any synthesis requires enough information to answer the following questions, but is not limited to them:

Policy Objectives

- What kinds of dynamic interactions between societies do our policies describe?
- What are the ranges of possible states and trends that meet policy objectives?
- What combinations of ways and means are off limits due to other policies?

Actor Types

- What level of abstraction is acceptable/sufficient in defining our actors?
    - Nation state
    - Official/ unofficial subgroups
    - Bureaucracies and organizations
- What are the most significant physical features of the actor?
    - Geostrategic context
    - Resources
    - Technical sophistication
    - Relative parity with other actors
    - Economy
    - Military capabilities

- What are the most significant cognitive features of the actor?
    - Grand narratives/stories/history
    - Cultures
    - Ideologies
    - Social norms
    - Stated policies & redlines
    - Alliances
- Phase
    - What are the current phase relationships between the actors in a bilateral sense?
    - Can the relationships between different actors be described by different simultaneous phases?
    - Do different combinations of actors imply different phase sensitivities?
- Threat
    - What are the relative strengths and weaknesses of various actors compared to one another in different areas
- Diplomatic
- Informational
- Military
- Economic
- International Context
    - What particular sensitivities are most relevant, given the specific historical, geographic, and actor specific contexts?

Agent-based models could be especially useful to analyze the linear aspects of specific scenarios (i.e. force flows, expendables tracking, basing suitability, force-on-force engagement results, etc.), and if the assumptions can be spelled out with the results, agent-based models might even be useful for collecting SME inputs to explore the social aspects of various scenario iterations.

Skill set for adapting CAS to the 5D Environment

While physical limitations of planning spaces and group dynamics may limit the practical size of the group that works immediately together, the more diverse the resident expertise is, the more likely it is that key variables and sensitivities will be identified as the group works through notional scenarios. Single members of the planning team will be able to fulfill several roles described below, depending on their training, operational background, and specific areas of expertise.

At the minimum, each planning team should have the following skill sets:

Specialized/Technical expertise:

- Tactical experts who have a detailed knowledge of the relative capabilities, strengths, weaknesses, and support requirements for all relevant friendly, hostile, and neutral military forces
- Regional specialists who are familiar with significant historical, cultural, economic, religious, and political issues, themes, and narratives in the areas of interest

- Personnel who are monitoring the current state of affairs of all relevant players, including physical disposition, public and private statements of intent & policy, and third party/public media interpretations of what is happing that may affect perceptions in key areas of interest
- Personnel who have detailed information on the formal and informal reward and punishment systems that drive societal behavior within areas of interest and can communicate which key individuals, groups, and processes control them, and by which means
- Legal experts who can brief regulatory constraints on freedom of action and assess the suitability of proposed courses of action

Organizational Expertise

- Operational level and organizational specialists who have a thorough understanding of the internal and external workings of military bureaucracy and processes and can communicate the opportunities and limitations involved in planning and executing operations through large organizations
- Representatives and liaisons from supported, supporting, and horizontally-connected stakeholder organizations outside of the military chain of command

Planning expertise

- Strategists who can assess the larger scope and longer-term implications and risk of proposed courses of action and who communicate different theories of causality and decision-making to the group
- Planning team leaders who are familiar with managing group planning processes who can steer the group discussions as needed but still keep in place group management techniques that minimize social biases like groupthink, mind-guards, the halo effect, etc.
- Planners and analysts who are familiar with the analytic tools that are available and will be used, who can relate their strengths, assumptions, limitations, and results to the larger group
- Briefers who are trained to use decision support tools and various presentation styles, who can communicate the findings of the group in a concise, timely manner using modes of communication that senior leaders and decision-makers can quickly assimilate

FURTHER RESOURCES

Allison, G., & Zelikow. P. (1999). *Essence of Decision: Explaining the Cuban Missile Crisis*. New York: Longman.

Bar-Yam, Y. (2004). *Making things work: Solving complex problems in a complex world*. United States of America: NECSI Knowledge Press.

Beaufre, A. (1967). *Strategy of action*. London: Faber and Faber.

Bousque, A. (2009). *The scientific way of war*. New York: Columbia University Press, 2009.

Dolman, E. C. (2005). *Pure strategy: Power and principle in the space and information age*. New York: Frank Cass.

Gardner, H. (2006). *Changing minds: The art and science of changing our own and other people's minds*. Boston: Harvard Business School Press.

Gharajedaghi, J. (2006). *Systems thinking: Managing chaos and complexity*. United States of America: Elsevier, Inc.

Holland, J. (1995). *Hidden order: How adaptation builds complexity*. Cambridge: Perseus Books.

Holland, J. (1998). *Emergence: From chaos to order*. New York: Basic Books Group.

Johnson, S. (2001). *Emergence*. New York: Scribner.

Johnson, S. (2010). *Where good ideas come from: The natural history of innovation*. New York: Riverhead Books.

Morgan, G. (2006). *Images of organization*. Thousand Oaks: Sage Publications, Inc..

Payne, K. B. (2011, Summer). Maintaining flexible and resilient capabilities for nuclear deterrence. *Strategic Studies Quarterly* 5, no. 2, pp. 13-29.

Ryan, A. (2009, Summer). The foundation for an adaptive approach: Insights from the sciences of complex systems. *Australian Army Journal* VI , pp. 69-90.

Sanders, T. I. (1998). *Strategic thinking and the new science*. New York: The Free Press.

Surowiecki, J. (2004). *The wisdom of crowds*. New York: Anchor Books.

Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of chaos*. New York: Simon and Schuster.

CONTENT ANALYSIS

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Time-series** : multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ❑ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

Content analysis is a technique that has been utilized since at least the 18th century and has grown and changed in frequency of use and method since the mid-20th century. The use of content analysis increased sharply when the U.S. Government (USG) funded research of mass communication during World War II. Various offices or departments (such as Library of Congress, Office of War Information, and the Foreign Broadcast Intelligence Service [later known as the Federal Communications Commission]) did content analyses of newspapers (see Lasswell, 1942), propaganda, magazines, news reels, and even comic strips (Berelson, 1952).

The most often cited definition of content analysis is Berelson's 1952 definition, "*content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication*" (italics original to author). Put very simply, the basic idea is for content analysis to be one process to get to meaning through people's use of symbols.

Content analysis has a history of being either a qualitative or quantitative method. However, it is generally also viewed as either not qualitative or quantitative *enough*. To this effect, Weber (1985) points out that there is no one right way of doing content analysis. A researcher has the ability and flexibility to adapt his or her procedure to appropriately address the specific research question or hypothesis.

METHODOLOGICAL OVERVIEW

WHAT IS CONTENT ANALYSIS?

Krippendorff (1980) broadly defined content analysis as a "research technique for making replicable and valid inferences from data to their context" (p. 21). These inferences could be about the sender of the message, the communication itself, or the receivers of the message.

Content analysis can be performed on any text. A **text** can be anything that communicates a message such as newspapers, books, television shows, book titles, signs, posters, record album cover art, commercials, etc. Content analysis is more than just a description of the content of a text. This method allows a researcher to make inferences about attitudes, opinions, and themes (from communication) that cannot be observed directly (from observation). Content analysis is basically a method to reduce data down to categories of content. Content can be classified either by human coders or through computer software. Because people interpret texts subjectively, the coding rules must be defined in such a way that the process could be reproducible by a different set of coders and receive the same results [reliability]. **Reliability** is defined as the "extent to which an experiment, test, or measuring procedure yields that same results on repeated trials" (Carmines & Zeller, 1979, p. 11). If the coding is automated, once the coding rules are defined, there is no need for inter-coder reliability. A computer, if the rules are defined correctly, will code a text the same way each time. If human coders are used, inter-coder reliability and reliability between coders must be calculated. Cohen's kappa coefficient is a standard statistical test of inter-coder reliability for categorical data. Also, Krippendorff's (2007) coefficient (α), developed for content analysis, is an inter-coder reliability test for any level of data (nominal, ordinal, interval, or ratio). Krippendorff (2004) suggests that a minimum coefficient, $\alpha>0.667$, be used as a tentative acceptance of inter-coder reliability and $\alpha>.800$, as a measure of good reliability. However, Krippendorff stresses that more conservative criteria should be set when the content analysis results will be used beyond "to merely support scholarly arguments," and instead will impact major decisions (2004).

Researchers can apply content analysis in both qualitative and quantitative ways. A **qualitative content analysis** is a subjective interpretation of a text through a process of coding and identifying themes or patterns (Elo & Kyngäs, 2008; Hseih & Shannon, 2005). **Quantitative content analysis** is basically coding text into explicit categories and then applying statistics to describe the data. Computer software can be used to speed up the coding process (see Hopkins & King, 2010 for an example of the preparation for automated content analysis).

Qualitative Content Analysis

Content analysis is a flexible form of analysis. There are three main types of qualitative content analysis: conventional, directed, and summative content analysis (Hsieh & Shannon, 2005). **Conventional content analysis** is best used when there is little theory or existing research on a topic. The codes and categories will emerge during study of the texts. After categories are identified, they can be described and defined. In **directed content analysis**, the content analysis is used to validate an existing theory. Existing research and theory help to inform the coding categories, interpretation, and discussion. **Summative content analysis** begins with a keywords search. A simple keyword frequency count is tabulated either by hand or through software. This quantitative step is used to explore word usage, not to infer meaning. These word counts are then used to facilitate more qualitative analysis such as looking at alternative meanings or entire text content.

Differences Between the Types of Qualitative Content Analysis*

Type of Qualitative Content Analysis	Study Begins With:	When Codes or Keywords are Defined in Process	Source of Codes or Keywords
Conventional	Observations	Codes defined *during* data analysis	Codes are derived from data
Directed	Theory	Codes defined *before* and *during* data analysis	Codes are derived from theory or previous research
Summative	Keywords	Keywords are indentified *before* and *during* data analysis	Keywords derived from literature or focus of researchers

* Table recreated from Hsieh & Shannon (2005)

Quantitative Content Analysis Techniques

Weber (1985) describes a few basic techniques to quantitatively analyze content in worded texts. Content (words, images) is explicitly put into categories and analyzed statistically. A few common techniques to quantify and analyze data are the key-word-in-context (KWIC) lists, category counts, and factor analysis. There are more techniques beyond these examples.

A **Key-Word-In-Context (KWIC)** list will show the word of interest as a pre-set amount of words that appear in the text surrounding that word. According to Weber (1985), KWIC lists allow the analyst to see the consistency or variation in meaning or usage of a particular word. Also, the KWIC list allows an analyst to systematically see how a word is used in a particular kind of phrase.

```
1   emed quite natural);  but when the Rabbit actually TOOK A WATCH OUT OF ITS WA
2   t a thousand times as large as the Rabbit, and had no  reason to be afraid of
3    `No, they're not,' said the White Rabbit, `and that's the  queerest thing ab
4   ging for apples, indeed!' said the Rabbit angrily.  `Here!  Come and help me
5   che!'      Alice watched the White Rabbit as he fumbled over the list,  feeli
6    `Did you say "What a pity!"?' the Rabbit asked.      `No, I didn't,' said Al
7   `She boxed the Queen's ears--' the Rabbit began.  Alice gave a  little scream
8   d the King.      On this the White Rabbit blew three blasts on the trumpet, a
9   ess,' said the King; and the White Rabbit  blew three blasts on the trumpet,
```

Figure 1. Example of a KWIC list.[3]

A **Word Frequency** list is a generated list of the most frequently used words in a text. Most software programs will omit frequent words such as *the*, *a*, *to*, and forms of *is*. Word frequency lists must be examined closely. Frequently used words do not necessarily reflect importance. All usage must be taken in consideration along with context. Word frequency lists could be a starting point for further inquiry.

Category Counts are a technique that is useful to highlight areas that receive more attention within a text. Category counts are simply counts of words that appear within particular categories based on the coding scheme. Category counts can be easily used to compare topics of focus between sources, between mediums, or across time.

Factory Analysis is a statistical technique using category counts to identify themes within a text. Each factor could represent a particular theme from texts. Factor analysis is a multivariate statistical approach and is beyond the scope of this paper.

Content Analysis Procedure

Weber (1985) and Krippendorff (1989) reviews the most common set of steps that make up a content analysis procedure. While content analysis is a flexible tool for analyzing documents and texts, this outline of steps ensures that the researcher is maintaining rigor and integrity.

1. **Design and Define.** In this first step, researchers must 1.) Define their context (the information of interest that is not directly observable); 2.) Look for available sources that may contain the relevant data; and 3.) Decide on an analytic construct that frames the inferred data-context relationship that is being analyzed. At this point, the researcher should define and describe what he or she plans to test. This will help decide whether using an inference-based interpretation of the data is valid and truly representative of what the researcher is testing.

[3] Image retrieved August 8, 2011 from http://www3.uva.es/martindelpozo/corpus/concordef.htm

2. **Unitizing.** The researcher is to define and identify a unitization (break down or classification into discrete units) scheme for the overall corpus and what is being studied. The sampling units must be defined, and should be representative of overall available texts.

   a. **Recording units** are the meaningful units to be analyzed. Common recording units are:
      i. Words- code individual words.
      ii. Sentences- code entire sentence.
      iii. Themes- unit of text that includes the perceived, a perceiver, action, and target of an action (see Holsti, 1969). Longer sentences may have more than one theme.
      iv. Paragraphs- code entire paragraph.
      v. Whole text- code an entire text. This is more useful when the text is short (e.g., newspaper headlines) (Weber, 1985).

      It is important to note that reliability tends to decrease when the recording units are larger, such as paragraphs or longer.

   b. Create the **coding scheme** or rules, which is the description of coding rules and categories that coders or software will use. The **categories** to be used to code the texts must be defined. Weber (1985) notes that researchers must decide if the categories are to be mutually exclusive and how that would affect later statistical analysis if units can be coding by more than one category. Also, when defining categories, researchers will have to decide on how narrow or broad the categories should be. Broader categories will have more possible entries than more narrowly defined ones.

3. **Test Sample**. A test of the coding scheme should be carried out to assess reliability. If computer reproducibility or inter-coder reliability is not satisfactory, the coding rules should be revised. This testing phase should be repeated until reliability is reached. A conservative metric of reliability would be a Krippendorff's alpha coefficient of $\alpha > 0.8$ or better (Krippendorff, 2004).

4. **Code All Text**. The remaining text can be coded once high coder reliability scores are reached or the computer coding software is performing reliably. Once the entire corpus has been coded, assess achieved reliability.

5. **Drawing Inferences**. Krippendorff (1989) describes this stage as an important but rarely obvious step. It is not enough to relate the coded data to the phenomena that the researcher is studying. Instead, the researcher must take the known information about the coded data to make inferences about the phenomena of interest.

6. **Validation**. While content analysis attempts to analyze the unobservable, at least some attempt should be made to validate or compare evidence based on the results of the content analysis. In one sense, true measurement validity should be analyzed. One example of measurement

validity is **convergent validity**, a comparison of how similar the outcome of the content analysis with the outcome of similar research. Krippendorff (1989) also discussed validation in the sense that content analysis should be done for 'valid' reasons, meaning that it is important for content analysis to be done for non-superfluous reasons. For example, Krippendorff asks "why would one want to extract military intelligence from enemy propaganda if the adversary's planned activities were already known" (1989, p. 407).

Advantages of Content Analysis

Content analysis is an adaptable method of analysis that still manages to be systematic and yet readily applicable to a wide variety of research questions and available data (Weber, 1985). Communication is an essential part of any kind of interaction and people purposefully or accidentally keep written (tangible or electronic) copies of our communication.

- Content analysis is an unobtrusive way to study the phenomenon or topic of interest.
- Content analysis does not require a researcher to engage with human participants which can be time-consuming and runs risks.
- Data for content analysis can come in many forms, including newspapers, television and film media, blogs, magazines, transcripts, textbooks, e-mail, etc. Given the reach of technology, a researcher has the potential to access a multitude of texts for analysis. If accessible, a researcher can look at texts across time, across medium, or across place of origin.
- Content analysis allows for the quantitative analysis of the relationships between "economic, social, political, and cultural change" (p. 10) that are reflected in the cultural indicators in our discourse.
- Content analysis can be used as part of an overall research plan (Weber, 1985).

Disadvantages and Limitations of Content Analysis

Content analysis is limited to the acquisition of appropriate texts. In some cases, obtaining texts is not difficult (e.g., first page headlines from the past one, two, ten, etc. years of the New York Times) but at other times, the texts may not be accessible or usable (e.g., because of security issues, in a foreign language, or even privacy concerns). The inability to obtain texts for analysis may end a research project. On the other hand, with widespread access to information as well as automated text-coding software, Weber (1985) also warned of the "dangers of mindless content analysis" (p. 69). Too much information runs the risk of everyone doing content analyses but not necessarily doing it well.

Content analysis as a method opens itself up for human interpretive or methodological error. Human coders add risk of inaccuracy and bias. The basis of content analysis is the identification and interpretation of symbols. Symbols have no one right definition or meaning and thus are always, on some level, subjective. The training of coders and having competent coders is thus an important aspect of this technique. Additionally, for coding to go as well as humanly possible, even with the best of coders, much success is dependent on the quality of the coding scheme.

REQUIREMENTS

DATA

Low (2)

A benefit of the content analysis is that data can come in a variety of forms and can depend on the topic of interest, the research question, and even the availability of texts. Content analysis can be performed on any text. A text can be anything that communicates a message such as newspapers, newspaper headlines, books, television shows, book titles, signs, posters, record album cover art, commercials, etc. Once a medium is chosen, the researcher can decide to look at a text at different levels. For example, if documents are used, the researcher could code individual words, sentences, themes, paragraphs, or even the entire document itself. The data requirements are intrinsically connected to the research question and focus. Parameters for the data collection cannot be generically appointed.

SET UP TIME

Average (3)

The time required to conduct content analysis will vary from project to project and will depend on a variety of factors. If texts are not readily available, acquiring materials to analyze can be a lengthy and time-consuming task.

EXECUTION TIME

Long (1)

Hand coding a corpus of documents will require more time than an automated coding procedure. For example, in an event coding project (identify event in piece of text such as a newspaper headline), human coders could identify approximately 40 events per day and the automated reference finder software could identify 2000 events per second (Schrodt, 2001).

SKILL SET/EXPERTISE

Specialized-average (1)

Content analysis does not require extensive knowledge of research techniques but a solid understanding of social science research techniques is beneficial. Qualitative and quantitative content analysis requires researchers to be able to formulate a research question or hypothesis, perform a thorough review of the literature, collect appropriate texts, conduct reliability assessments of a coding scheme, train coders, and maintain data. Quantitative content analysis will require understanding of statistical methods, database management, and if used, ability to use coding software.

Coders must be appropriately trained in order to competently contribute but do not necessarily have to be subject matter experts (SMEs). The researcher has the responsibility to create a reliable and accurate coding scheme and then create a thorough coding manual and training approach for coders.

TOOLS

Specialized (2)

Today, there are many options regarding content analysis software. Reviews of software according to the need of the researcher would be most useful (Hill, 2008; Lowe, n.d.). Familiar software includes LIWC (Linguistic Inquiry and Word Count) and Concordance. In some cases, software choice will depend on operating system (Windows, Mac, or Linux). Statistical software, such as SPSS or SAS will be required if a quantitative approach is used.

COST

$$ (9)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

The 5D framework asks three general questions: 1) How effective is current US force posture for achieving policy objective?; 2) What would be the optimal force posture to achieve a specific policy objective?; and 3) What strategy is optimal to achieve the objective?

1. *How effective is current US force posture for achieving policy objective?*

   Our current force posture does not appear to draw on content analysis. For instance, in the Nuclear Terrorism Deterrence project several years ago, no content analysis was done. The closest work seems to be entity extraction as done in classified contexts. Entity extraction is basically a word search and identification—but not true content analysis.

   One effort that comes close to content/narrative analysis is The Islamic Imagery Project done at the Combating Terrorism Center at West Point. It is an analysis of pictorial imagery used by Jihadists. The significance of particular images (e.g., animals, colors, people, and other symbols, like hands in prayer) are discussed within the context of Jihadi images, not just images associated with Islam or particular countries or groups.

2. *What would be the optimal force posture to achieve a specific policy objective?*

   Written or otherwise recorded human communication holds the possibility of extensive analysis. Historical and/or current communication can be examined repeatedly as new data, technology, or theories emerge. People will continually people betray their sentiments and intentions through the rhetoric they use. Therefore, maintaining a robust national capability for the analysis of the content of communication, for dealing with adversaries and allies, states and non-states, is necessary. These capabilities have likely homes in the Department of State, the non-military intelligence community, and the military intelligence, planning communities and civil affairs.

Fielded units engaged with communication with adversaries and host populations/governments should have some basic familiarity with content analysis. They may not have the resources and time in theatre to conduct such analyses, but they at least need to be aware of the importance of examining communication for underlying sentiment and narrative and be aware of what kinds of information they could gather and what kinds of analyses that could inform their own operations.

Large scale content analysis requires time and technical resources. Reach back cells should be maintained in order to supply such analyses. Some cells already exist, such as with Human Terrain System regional reach back cells in the US, who support the research of fielded human terrain teams.

3. *What strategy is optimal to achieve the objective?*

The first step is assessment of current military and intelligence community (IC) capabilities for content analysis. Who does it now? Where? What are their challenges and needs? A gap analysis should identify who should have these capabilities and how their personnel could be trained or new personnel hired to fill those gaps.

Achieving a robust capability for content analysis in the military and IC requires government funded R&D of the nation's brain trust in academia and industry. The research should support both theoretical advances in understanding human communication and behavior, as well as methodological advances in analysis of texts and narratives. Milestones for success in this research need to be defined, and as specific projects mature, clear pathways to transition to military training and operations must be defined, including the identification of the military and non-military specialties where these capabilities are most needed and best used, and plans for how new methods, techniques and procedures would be trained, accredited and incorporated into military and intelligence community business.

Another area of research that should be supported should be the correlation of what people say, what they really mean, and deception, since deception is a fundamental pillar of any influence operation and must be considered when analyzing an adversary's rhetoric.

FURTHER RESOURCES

Basic References

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Krippendorff, K., & Bock, M. A. (Eds.) (2009). *The content analysis reader*. Thousand Oaks, CA: Sage Publications.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Weber, R. P. (1985). *Basic content analysis*. Beverly Hills, CA: Sage Publications.

Application Towards Nuclear Strategy

Brody, R. A. (1963). Some systemic effects of the spread of nuclear weapons technology: A study through simulation of a multi-nuclear future. *Journal of Conflict Resolution, 7*(4), 663-753.

Rehm, B. D. (2006). *French nuclear strategy in an age of terrorism*. (Master's thesis, Naval Postgraduate School). Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA462671&Location=U2&doc=GetTRDoc.pdf

WORKS CITED

Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: The Free Press Publishers.

Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment. Vol. 17*. Thousand Oaks, CA: Sage Publications.

Combating Terrorism Center at West Point (2006, March). *The Islamic imagery project: Visual motifs in Jihadi internet propaganda*. Retrieved from http://www.ctc.usma.edu/posts/the-islamic-imagery-project

Elo, S. & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107-115. doi: 10.1111/j.1365-2648.2007.04569.x

Hill, K. G. (2008). Software for content analysis: Links to external sites. Retrieved from http://courses.washington.edu/socw580/contentsoftware.shtml

Holsti, O. (1969). Content analysis for the social sciences and humanities. Menlo Park, CA: Addison Wesely.

Hopkins, D. J. & King, G. (2010, January). A method of automated nonparametric content analysis for social science. *American Journal of Political Science, 54*(1), 229-247.

Hsieh, H., & Shannon, S. E. (2005, November). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Krippendorff, K. (1989). Content analysis." In E. Barnouw et al. (Ed.) *International Encyclopedia of Communication*, Vol. 1. New York: Oxford University Press.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research, 30*(3), 411-433.

Krippendorff, K. (2007, June 1). Computing Krippendorff's alpha-reliability. Retrieved from http://www.asc.upenn.edu/usr/krippendorff/webreliability2.pdf

Lasswell, H. D. (1942). *Analyzing the content of mass communication: A brief introduction.* Library of Congress, Experimental Division for Study of War-Time Communications, Document no. 11. Washington, D. C.

Lowe, W. (n.d.). Software for content analysis: A review. Retrieved from http://dev.wcfia.harvard.edu/sites/default/files/identity_content_analysis.pdf

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Post, J., Sprinzak, E., & Denny, L. M. (2003, Spring). The terrorists in their own words: Interviews with 35 incarcerated Middle Eastern terrorists. *Terrorism and Political Violence, 15*(1), 171-184.

Schrodt, P. A. (2001, February). *Automated coding of international event data using sparse parsing techniques*. Paper presented at the annual meeting of the International Studies Association, Chicago, IL.

Weber, R. P. (1985). *Basic content analysis*. Beverly Hills, CA: Sage Publications.

CROWDSOURCING

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☐ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☐ **Sub-national/organization group** (e.g., Pakistani military)
- ☐ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☐ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ☐ **Quantitative**
- ☑ **Qualitative**
- ☐ **Quantitative and qualitative**

INTRODUCTION

SME crowdsourcing is an analytic technique to gather socio-cultural insights on topics of interest to U.S. government analysts, planners, and decision-makers. Socio-cultural insights can be especially useful for understanding the motivations, interpreting the actions, and ultimately influencing the behavior of foreign actors. By illuminating the complex array of socio-cultural lenses through which real-world decision-makers see the world, this analysis can help U.S. government analysts and planners step into those foreign mindsets, understand where there may be discord or unanimity on key issues, and imagine how these fault lines may shape interpretations of future actions by the U.S. or others in the region.

SME crowdsourcing represents a powerful alternative analytic approach that can complement rational actor and strategic decision-making models that are often applied to questions related to deterrence, assurance, and non-proliferation. For example, rational actor models may help answer a question such as, "What is Japan's decision calculus regarding nuclear proliferation, based on prior behavior, known constraints, and how decision-makers assign values to risks and benefits?" While this is a useful question for understanding how "rational" actors will act or react, given a set of conditions or specific scenarios, the rational actor approach may miss the full set of nuanced beliefs sets that govern decision-making among foreign actors and influencers. SME crowdsourcing is not intended to replace existing rational actor and decision-making models. Rather, the socio-cultural insights it produces can provide critical context for better formulating the nuclear planning problem and for applying those other techniques.

METHODOLOGICAL OVERVIEW

SME crowdsourcing is used to gather and synthesize perspectives from a wide array of experts and thought-leaders to shed light on complex analytical problems and to offer alternative (and sometimes competing) perspectives.

SME crowdsourcing can operate on multiple levels of analysis: on both a country-specific level as well as on a regional level that synthesizes crosscutting insights. This technique, as employed by Monitor 360, relies on qualitative research and analysis rather than a heavy quantitative component. However, this analysis can be layered upon quantitative models or survey methods depending on the needs of analysts and planners. While other SME elicitation techniques may focus on gathering expert opinion on a narrowly-defined set of questions, the technique employed by Monitor 360 gathers insights on multiple levels of granularity, thereby providing U.S. government analysts and planners with several windows into the foreign mindsets of interest.

This technique can be used to address questions for which U.S. Government decision-makers, analysts, and planners seek to better understand how foreign actors view an issue, as well as to better anticipate how they are likely to act or react in the future. SME Crowdsourcing for socio-cultural insights can be used for questions that are tied to current or potential “influence actions” abroad, which would rely on a rich understanding of the groups who are targets of influence.

SME crowdsourcing can also be useful to wargame planners by providing socio-cultural inputs into pre-game planning and “scripting” of potential moves. In addition, it can serve as a resource for game players to more effectively get into the mindsets of the foreign actors being portrayed in the game. Finally, this analysis may also prove useful for post-game analysis, as the insights from SME crowdsourcing can be compared with the results from the game itself.

HOW IS THIS DONE?

The outputs of this analysis are broken out into three key elements.

1. <u>Key Segments:</u> The major clusters of opinion on regional security issues among decision-makers and influencers in a given country
2. <u>Master Narratives and Sub-Narratives:</u> The “timeless” cultural and historical lenses through which influencers and decision-makers in the region will interpret events and policies today and in the future

3. Implications for Wargame Planners: A set of hypothetical moves and potential interpretations of those moves by different Key Segments, drawing on the Master Narratives and Sub-Narratives surfaced through Monitor 360 analysis

SME crowdsourcing for socio-cultural insights is a process that involves three major steps in order to derive the three elements outlined above:

1. Initial SME interviews and open-source research: First, the team conducts an initial set of interviews to sketch out the landscape of opinion on relevant regional security issues in the region of interest. These initial SME interviews are used to develop hypotheses about the relevant key segments and narratives that may exist in each of the countries of interest. These initial hypotheses are supported by a review of open-source research materials, including primary sources, such as domestic media and historical texts, and secondary sources, such as foreign media and academic publications.

2. In-depth interviews with regional SMEs: Next, the team conducts a set of "deeper-dive" interviews with SMEs that include academic and policy experts, former officials, thought leaders, and influencers aimed at developing hypotheses about the clusters of key beliefs and opinions in the region of interest. Interviewees are selected based on their ability to both dive deeply into country nuances, as well as their ability to make connections at the regional level across countries. These deeper-dive SME interviews are used in order to gain greater insight into how the narratives are understood at the local level and who the major players are in each country that subscribe to them. This deep-dive phase allowed the team to gather new "narrative seeds" (content that may inform development of new narratives), as well as to test initial hypotheses about how narratives map to key segments.

3. "Validation interviews" with SMEs to test and refine these hypotheses and to develop the underlying narratives that capture the various perspectives in each country.

4. Writing Final Outputs: Finally, the team writes up the Key Segments and the Master Narratives and Sub-Narratives themselves. The writing phase involves the construction of several common "elements" of a narrative, as well as the synthesis of these elements into cohesive whole. The team's writing activities include outlining the plot, detailing the historical and present-day manifestation of the narrative, identifying the major "characters" (both protagonists and antagonists), and ultimately surfacing the final "call to action" which appears at the end of each Master Narrative and Sub-Narrative.

REQUIREMENTS

DATA

Average (3)

The type of data collected for this method is qualitative, taking the form of several clusters of opinion and perspective on issues of relevance to US planners and analysts, which are then synthesized and written up into a final report format. The data is categorized into various "buckets" based upon the specifics of the SME elicitation effort and the key focus areas for each of the SME interviews. Open-source research is a key component that can complement and provide context for the insights gained from SME interviews. Conducting open-source research requires a team that is skilled in research methods, as well as open access to a wide variety of web-based materials, books, reports, and foreign resources (such as reports from a specific wing of a foreign government or a military journal).

SET UP TIME

Short (2)

The time required for this analysis can vary depending on the specific questions being asked, as well as the geographic scope of the analysis.

EXECUTION TIME

Average (0.5)

In the CANS effort, this analysis took 4-5 months to compile the analysis and deliver findings focused on three countries, which comprises a medium-sized crowdsourcing effort compared to others that have been done for the military. An effort that examines more countries or regions, or that does a deeper dive on more specific issues, may take longer and, thus, comprise a “larger” effort in scope.

SKILL SET/EXPERTISE

Diverse-advanced (4)

In order to use this method, the team conducting the analysis must have access to a diverse network of SMEs, as well as a capacity for synthesizing the outputs of SME crowdsourcing into findings that are useful to US government planners and analysts. These key issues need to be taken into account to carefully design SME interview plans that would extract opinions and perspectives on these key issues. Team members working on SME crowdsourcing effort have demonstrated previous experience with complex analytic and geostrategic problems, as well as previous experience in research design and writing.

- Access to SMEs
- Mastery of SME elicitation skills and techniques
- Mastery of open source research skills Mastery of analytic skills to gather and synthesize insights
- Writing skills to convey complex socio-cultural concepts in user-friendly form aimed at decision-makers, analysts, and planners

TOOLS

Average (1)

COST

$$ (10.5)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Socio-cultural insights from SME crowdsourcing can help address U.S. nuclear policy planning questions by:

- Surfacing the motivations, beliefs, and attitudes of key foreign decision-makers and influencers
- Offering a lens through which to interpret past and future actions or key decisions by foreign actors
- Providing critical context for formulating influence actions aimed at key actors, groups, and organizations

By illuminating the complex array of socio-cultural lenses through which real-world decision-makers and influencers see the world, this analysis can help U.S. government analysts and planners **step into those foreign mindsets, understand where there may be discord or unanimity on key issues, and imagine how these fault lines may shape interpretations of future actions by the U.S. or others in the region**.

Users of the 5D framework can utilize socio-cultural insights in two primary ways:

1. **As Additional Context for Formulating Nuclear Policy Planning Problems:** Sociocultural insights can help provide additional context for questions related to assurance, deterrence, and proliferation. These insights provide nuclear policy planners with a set of inputs that can shed light on mindsets of foreign actors who are at the heart of planning and the formulation of U.S. policy objectives.

   For example, an assurance question might be: *How can the U.S. assure South Korea of the strength of the U.S. commitment?*

   SME crowdsourcing revealed two key segments in South Korea, with divergent views on the relationship with the U.S. The insight that these two segments in South Korea hold very different views on the importance and effectiveness of the U.S. security umbrella can help analysts articulate and better assess questions such as:

   a. Who, exactly, is currently being “assured” in South Korea about the strength of U.S. commitment?

   b. How might each of these segments require different strategies to be assured, and what would these strategies mean relative to current nuclear policy planning agendas?

2. **As a Complement to Other Analytic Techniques:** SME crowdsourcing for socio-cultural insights can complement other analytic techniques after the "problem formation" stage of 5-D framework, such as wargaming and modeling/simulation.

There are at least **three opportunities for integrating** socio-cultural insights from SME crowdsourcing into other analytic techniques, such as modeling and wargaming:

- <u>As a tool for scoping and design;</u> For example, wargame planners may select moves or players/teams for a game based on different nuances and viewpoints held by Key Segments. Using these insights in the design phase may help planners consider the tensions and fault lines worth testing in the game.

- <u>As a real-time feedback loop</u>: For example, modeling and simulation teams may use socio-cultural insights to test preliminary results and to re-focus or re-scope before completing subsequent phases. The insights could provide modeling teams with a "feedback loop" against which to calibrate their findings as they run their models, in order to improve their overall results.

- <u>As an alternative approach for comparative results</u>: For example, socio-cultural insights can be useful for post-game analysis in order to compare insights from a wargame and to "wind tunnel" the moves used in a wargame after it ends. These insights may also be used to generate plans for future wargame efforts.

FURTHER RESOURCES

An extensive compilation of further resources used by Monitor 360 is contained in the report of findings, delivered on June 1, 2011.

EXPERIMENTATION

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☐ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☐ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☐ **Single nation-state or non-state actor Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☐ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ☐ **Quantitative**
- ☐ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

Experimentation is a methodology most commonly associated with the natural sciences. It is widely used in psychology, and its advantages for addressing certain types of questions are increasingly recognized in economics and political science. One of the central advantages of experimentation over other methods is that it enables analysts to test causal relationships, rather than having to make causal inferences from statistical correlations or qualitative analyses. Experiments are also used in nonscientific human inquiry. In cooking, for example, we may make the same basic dish multiple times, adjusting elements (seasoning, ingredients, cooking time) each time until we achieve the "best" version. From the time we are small children, we experiment frequently in our attempts to develop a generalized understanding of our world.

METHODOLOGICAL OVERVIEW

WHAT IS EXPERIMENTATION?

Experimentation attempts to test causal relationship by demonstrating that changes in one variable are directly responsible for changes in another variable. For example, we are probably all familiar with the **classical experimental design** used in drug trials:

> *To test the effectiveness of a new cholesterol-lowering drug you would choose two groups of people (sharing the same basic characteristics: age, sex, health, etc). Before starting the experiment, the*

cholesterol level of each participant would be measured. One group (control) would be given a placebo to take for the duration of the experiment, while the other group would be given the new drug (experimental). At the end of the experimental period, all participants' cholesterol levels would be measured again. If participants in the experimental group (who took the drug) show significantly greater decreases in their cholesterol levels compared to those in the control group, then the experimental findings support the claim that the new drug can lower cholesterol levels.

The classical experimental design is most common to the natural sciences and involves assessing the effects of an independent variable on a dependent variable. The participants in the experiment are divided into an experimental group (who receive the stimulus) and the control group (who do not). The researcher compares what happens when the stimulus is present to what happens when it is not by measuring the dependent variable prior to and after the experiment. If the degree of change in the dependent variable differs between the control and the experimental group, then the stimulus (as the only difference between the two groups) is known to be the cause.

The classical experimental design works well for many research questions in the natural sciences. However, it is not necessarily an ideal design for experimental research in the social sciences when in many cases we are interested in how different types of stimuli change responses, rather than comparing stimulus to the lack of stimulus. The choice of experimental design is determined by: 1)The research question, 2) how the independent variable is measured, and 3) whether subjects will be assigned to only one treatment condition (**between-subjects design**) or each subject will participate in all treatment conditions (**within-subjects design**).

A simple **post-test design** is similar to the classical experimental design. In this design there are two randomly-assigned groups, but there is no pre-test component. The experimental group is exposed to a treatment and the control group is not exposed to the treatment. Following treatment, the dependent variable is measured for each group. As random assignment of subject ensures that the only difference between the two groups is the introduction of the treatment in the experimental group, any difference in the dependent variable between the experimental group and the control group must be caused by the treatment.

For example, if we wanted to test the hypothesis that reading a newspaper article about a Supreme Court decision makes people better informed of the decision, we would randomly assign our subjects to two groups, one group that would read a newspaper article about a Supreme Court decision (the experimental group) and one group that would not read the newspaper article (the control group). We would then measure how well informed members of the two groups were after the experimental group had read the newspaper article. You will notice that there was no pre-test of the subjects' level of information prior to the debate. By randomly assigning subjects to the two groups, we assumed that each group's level of information was equivalent prior to the treatment. In other words, the explanation for any difference in level of information between the groups was the treatment, reading the newspaper article. This is normally a safe assumption if you have access to a large number of subjects and the subjects are randomly assigned to the two groups. However, if the group assignment is not truly random or if your sample size is small, the post-treatment differences between the two groups may be the result of pre-treatment differences rather than the result of the independent variable.

A **Solomon Four-Group** experimental design is a combination of a classical experiment (pre-test/post-test) and a simple post-test. Once again, subjects are randomly selected and, in this case, assigned to one of four groups. Group one receives a pre-test, treatment, and post-test (classical experiment experimental group). Group two receives only the pre-test and post-test (classical experiment control group). Group three receives the treatment and a post-test (simple post-test experimental group), and group four receives only the post-test (simple post-test control group).

For example, if we wished to test the same hypothesis as above, that reading a newspaper article about a Supreme Court decision made people better informed of the decision, in the Solomon Four-Group design, we would randomly assign our subjects to four groups. Prior knowledge about the Supreme Court decision would be measured in groups one and two (pre-test). Following the pre-test, groups one and three would receive the treatment (the newspaper article about the Supreme Court decision). Following the treatment, all four groups would receive the post-test. Remember, group four is only receiving the post-test.

This design allows us to test not only for the differences in the measurement of the dependent variable (knowledge of the Supreme Court decision), but for the effects of the pre-test and the effects of the treatment. The major drawback to this experimental design is the necessity for four groups and, thus, the number of subjects required.

A **multi-group design** provides the opportunity to compare different levels of the independent variable. This type of design is useful if the independent variable can assume several values. Multi-group designs may involve a pre-test/post-test or may simply involve a post-test.

In our previous examples, we were testing the hypothesis that reading a newspaper article about a Supreme Court decision made people better informed of the decision. What if, rather than wanting to merely assess whether people become better informed by reading a newspaper article about a Supreme Court decision, we wished to determine whether the bias of a newspaper article influenced whether a person agreed with a Supreme Court decision? Are people influenced more by negative media bias or positive media bias? Our hypothesis that the direction of media bias, positive or negative, will influence agreement with Supreme Court decisions presents an independent variable (direction of media bias) that can assume several values (positive, negative, or neutral). Thus, this hypothesis cannot be addressed by a classical experimental design, a simple post-test design, or a Solomon Four-Group design which all require an independent variable that is either present or absent (two values). In our new hypothesis, our independent variable (direction of media bias) can assume several values; the bias could be negative, positive or neutral. To test this hypothesis, we would need a total of three randomly-assigned groups – one to test each level of bias (positive or negative) and one control group (neutral, or no bias).

As with the Solomon Four Group design, the multi-group design requires more subjects than the pre-test/post-test or post-test only experimental designs. The more values associated with your independent variable, the more groups necessary and, thus, the more subjects necessary.
While these examples are only a sampling of the types of experimental designs that could be used to test various hypotheses[4], all experimental designs contain the following basic elements:

1. Manipulation. The researcher manipulates one variable by changing its value to create a set of two or more treatment conditions.
2. Measurement. A second variable is measured to obtain a set of scores in each treatment condition.
3. Control. All other possible sources of variation are controlled to be sure that they do not influence the experimental results.
4. Comparison. The scores in one treatment condition are compared with the scores in another treatment condition. Consistent differences between treatments are evidence that the manipulation has caused changes in the scores.

WHY USE EXPERIMENTATION?

Much like formal models, experiments are designed primarily to test hypotheses deduced from a given theory and model. Experiments do not test reality; rather, they test theories we have regarding human behavior in specific instances. In cases where the experiment is an appropriate representation of the theory, the findings merely support the logic of the theory.[5] How relevant the findings are to our understanding of the real world phenomena is a function of how accurately our theory captures the salient aspects of that phenomenon. Because experiments do not require real world data, they can be used to explore the consequences of controlled counterfactual scenarios, "what ifs," that are derived from more loosely defined theories.[6]

Finally, experimental methodology can be a very effective way of beginning to understand complex phenomena. Because the researcher can control what factors of interest (variables) change and which remain constant, he can break down complex relationships and explore particular theoretical links in the presence or absence of other factors.[7] "Careful experimental research design frequently helps sort out competing hypotheses more effectively than does trying to find the precise combination of variables in the field.[8]" That is, when faced with the common dilemma of a complex

---

[4] For a more thorough discussion of the various types of experimental designs, see Campbell and Stanley (1963).
[5] Geva, N. and M. Skorick. 2001. "The Cognitive Calculus of Responsiveness to Positive and Negative Feedback in Sequential Decisions." Paper prepared for the symposium on "*Responding to Negative Feedback in Foreign Policy Decision Making"* May 18-20, 2001. Center for Presidential Studies, George Bush School of Government and Public Service, Texas A&M University.
[6] Mook, D.G. 1983. "In Defense of External Validity." *American Psychologist* 38:379-388.
[7] McDermott, R. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10(4):325-42.
[8] Ostrom. E. 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action Presidential Address, *American Political Science Association, 1997." American Political Science Review* 92(1):1-22.

problem with a myriad of possible contributing factors and little data to examine, a careful experimental design can be more effective in refining our understanding of the underlying relationships than statistical or case study techniques.

It is important to note, however, that the distance between the experimental and the real world presents hazards, as well as opportunities. Therefore, the transition from theory to experimental design needs to be undertaken with great care. It is crucial to remember that a good experimental design is not one which replicates reality, but one in which the researcher can isolate causation, test theories, and generate hypotheses. As such, any experiment should be seen as a first, very small step; isolating the most basic component of both theories and determining their effects on the variable of interest. Once this basic relationship is established, additional complicating factors can be added and comparisons drawn between these controlled experimental results and those case and statistical studies drawing on real-world data.

How to Create an Experiment

The general purpose of experimental research is to test a hypothesized causal relationship, and it begins by defining a research question. Based on theory, research questions are defined and hypotheses are constructed that specify the causal relationship you are interested in testing. The emphasis in experimental research is on the development of specific **causal hypotheses,** hypotheses that reflect how the change in one variable results in a change in another variable.

Based on the specific hypothesis, the researcher must:

- Identify the independent and dependent variables. The independent and dependent variables determine what type of subject design you must use (between-subjects or within-subjects) and what type of experimental design is appropriate to address your hypothesis.
- Design the experimental scenario that will form the context for the experimental treatment of the independent variables
- Develop pre-test and post-test measurements (if applicable) of all variables
- Determine how subjects will be selected and how they will be assigned to the various experimental conditions (which treatment, if any, they will receive).
- Check for problems with internal, external, or construct validity.

To understand the experimental process, consider our earlier example of media bias and Supreme Court decisions. Our research question, "Does media bias influence agreement with Supreme Court decisions?" presents the following hypothesis: the direction of media bias, positive or negative, influences agreement with Supreme Court decisions, a hypothesis which can be tested experimentally.

Independent and Dependent Variables

An experiment examines the effect of an independent variable on a dependent variable. As noted earlier, the first requirement in experimentation is that of manipulation. In experimentation, the

independent variable is the variable manipulated by the researcher. The independent variable usually consists of two or more treatment conditions to which participants are exposed. In our example, we have hypothesized that the direction of media bias influences agreement with Supreme Court decisions. Our independent variable, the variable we will manipulate, is "media bias." This variable consists of three treatment conditions: positive bias, negative bias, and neutral (no) bias.

The independent variable also determines the structure of your experiment, how your experiment is going to be set up. It is the independent variable that guides the type of experiment you can do. In this example, our independent variable (media bias) has three values (negative, positive, neutral), meaning we are not merely looking for the presence or absence of the variable. Thus, a pre-test/post-test, simple post-test, or Solomon Four-Group design would not be appropriate, as they are merely testing for the presence or absence of the independent variable. In this case, we would need to use a multi-group design that accommodates the multiple values of our independent variable. In addition, our hypothesis requires us to use a between-groups design, a design where subjects are randomly assigned to one group, rather than a within-group design where subjects are assigned to multiple groups.

The **dependent variable** is the variable that is observed for changes to assess the effects of manipulating the independent variable. The dependent variable is typically a behavior or a response measured in each treatment condition. In terms of cause-and-effect, the independent variable is the cause, and the dependent variable is the effect. In our example, the dependent variable, the variable we are trying to explain, is "agreement with a Supreme Court decision." One of the advantages of an experimental design is that multiple dependent variables can be measured. For example, in this design we might also be interested in *why* the subject agreed with the Supreme Court decision or how strongly they felt about the issue.

The independent and dependent variables appropriate to experimentation are nearly limitless. It should be noted that a given variable might serve as an independent variable in one experiment and as a dependent variable in another. It is essential that both independent and dependent variables be operationally defined for the purposes of experimentation. Such operational definitions might involve a variety of observation methods. Experimentation requires specific and standardized measurements and observations.

Treatment Condition

A **treatment condition** is a situation or environment characterized by one specific value of the manipulated (independent) variable. Treatment conditions are chosen to focus on particular features of the testing environment. These conditions are administered to subjects in such a way that observed differences in behavior can be unambiguously attributed to critical differences among various treatment conditions. In our example, the manipulated (independent) variable is "media bias." To test the effects of media bias on agreement with Supreme Court decisions, we will develop three treatment conditions (three groups), positive bias, negative bias, and neutral bias.

We will manipulate this variable by asking subjects in one group to read a positively-biased newspaper article about a Supreme Court decision, a second group to read a negatively-biased newspaper article about the same Supreme Court decision, and a third group to read a newspaper article about the same Supreme Court decision that has no clear bias.

Pre-Tests and Post-Tests

As noted earlier, measurement is one of the four basic components of experimental design. In the simplest experimental design, subjects are measured on a dependent variable (**pre-tested**), exposed to a stimulus (the **treatment**) that represents the independent variable, and then re-measured on the dependent variable (**post-tested**). Differences noted between the first and second measurements on the dependent variable are then attributed to the influence of the independent variable.

In our example, we ask subjects to complete a pre-test questionnaire which would include a variety of questions, including their position on a series of policy issues, one of which is related to the Supreme Court case used in the experiment (pre-test). Subjects are then asked to read a summary of a recent Supreme Court decision and a newspaper report discussing that decision (treatment) and then respond to a series of questions regarding the case and their opinion on the particular issue (post-test).

Experimental and Control Groups

Experimental research also requires the ability to compare, specifically to be able to compare the scores in one treatment condition to the scores in another treatment condition. Thus, social science experiments seldom involve only the observation of an **experimental group** to which a stimulus has been administered. Researchers also observe a **control group** to which the experimental stimulus has not been administered. Using a control group allows the researchers to control for the effects of the experiment itself.

In our example, subjects would be randomly assigned to one of three groups, the negatively-biased article group (experimental group), the positively-biased article group (experimental group), or the neutral article group (control group). Because we are using a multi-group design, we can compare not only the experimental groups to the control group, but compare the negatively-biased experimental group to the positively-biased experimental group, increasing our understanding of the effect of media bias on agreement with Supreme Court decisions. Does a negative media bias influence agreement (or lack of agreement) more than a positive media bias?

Selecting Subjects

The fundamental rule of subject selection for experimentation is the comparability of the experimental and control groups. Ideally, the control group represents what the experimental group would have been like if it had not been exposed to the experimental stimulus. It is essential, therefore, that the experimental and control groups be as similar as possible.

It is important to note at this point the difference in requirements between survey and experimental research. In survey research, we are typically interested in understanding the opinions or behavior of a specific population. Survey results are only generalizable to the populations they represent. In experimentation, however, we are interested in testing a specific theoretical expectation, (in this example, the idea that media bias can influence support for a Supreme Court ruling). Unless there are any theoretical reasons to expect one segment of a population to be effected differently by the independent variable, then it is not necessary to have a set of experimental subjects that are representative of the population. For example, in the Supreme Court example, we had no theoretical reason to expect age, education, income, race, or any other individual characteristic to influence how the independent variable of interest (media bias) worked. Therefore, we could use a population of convenience, in this case college students, without needing to worry about the generalizability of our findings. As Mook states, in most experiments "[w]e are not *making* generalizations, but *testing* them." Therefore, "As to the sample: Am I (or is she or he whose work I am evaluating) trying to estimate from sample characteristics the characteristics of some population? Or, am I trying to draw conclusions, not about a population, but about a theory that specifies what *these* subjects ought to do" (1983: 386). If the former, sample selection matters; if the latter, it is only necessary to ensure that there is no systematic difference between the experimental groups. This can be achieved through random assignment into treatment groups.

Randomization

Randomization is another central feature of control in experimental design. In experimentation, randomization is most commonly achieved by assigning subjects into treatment conditions without reference to any individual differences or characteristics. The most important characteristic of randomization is that it produces experimental and control groups that are statistically equivalent. By doing so, it eliminates the possibility that the results of the experiment are due to individual characteristics of the subjects, rather than the experimental manipulation of the independent variables. If subjects are assigned to treatment groups through a random process, the assignment process is said to be unbiased and the groups are equivalent.

Once a group of subjects has been recruited, they are randomly assigned to either the control group or the experimental group. This may be accomplished in a number of ways. The subjects could be numbered and a random-number table used to assign each subject to a group. Or, you might assign all odd-numbered subjects to the control group and all even-numbered subjects to the experimental group.

In our example, using college students, we could assign each participant a number from one through three. The students with number 1 would receive the positively-biased media report, the students with number 2 would receive the negatively-biased media report, and the students with number 3 would receive the neutral media report.

Manipulation Checks

In an experiment, the researcher always manipulates (controls the value of) the independent variable. Although this manipulation and its results are obvious to the researcher, occasionally there is some question about the effect of the manipulation on the participants; are the participants aware of the manipulation and, if so, have they interpreted it in the manner intended? Researchers often include a manipulation check as part of the study in order to directly measure whether the independent variable had the intended effect on the participant.

There are two ways to check the manipulation. A manipulation check may be an explicit measure of the independent variable. For example, if you are measuring the effects of mood on the participant's performance (mood being the independent, or manipulated, variable), you might ask participants to indicate their mood on a measurement scale to ensure that the mood you were intending to achieve was actually achieved by participants.
You can also check the manipulation through the use of an exit questionnaire where, in addition to asking questions, such as whether the participant enjoyed participating, whether they were bored, etc., you can ask a specific question that addresses the manipulation. For example, if, as above, you were manipulating mood, you could ask the participant if they experienced a change in mood during the experiment.

In our example, we would want to assess whether subjects perceived the media reports accurately. For example, did they understand that the media report was negatively biased? This could be evaluated by including a question such as: "Which of the following statements reflects your understanding of the news coverage of the Supreme Court Decision?" Subjects would have the option of choosing one of the following: "The coverage of the Court's decision was unbiased," "The coverage was critical of the Court's decision," or "The coverage was supportive of the Court's decision." This manipulation check would assess whether the subject was aware of the bias in the media report.

Controlling for Threats to Internal Validity

Threats to **internal validity** compromise the researcher's ability to say whether a relationship exists between the independent and dependent variables. Did the experimental treatment make the difference, or was the difference caused by some other variable? When internal validity is high, differences between groups can be confidently attributed to the treatment, ruling out rival hypotheses. The following can affect the internal validity of an experiment:

History: Historical events may occur during the course of the experiment that confound the results. In our example, as the pre-test/treatment/post-test were all being conducted at the same time, there is little concern that historical events could affect our results.

Maturation: In a long-term experiment, the fact that subjects grow older may have an effect. In shorter experiments, subjects may become tired, sleepy, bored, or hungry, or change in other ways that affect their behavior in the experiment. In our example, as our experiment takes a very short period of time, there is little concern for a maturation effect.

Pre-Tests: The process of testing and re-testing may sometimes influence people's behavior and influence the results. In our example, subjects might make the connection between the pre-test questions about policy issues and the material they are asked to read about the Supreme Court decision. This might influence their response to the post-test questions.

Measuring Instruments: Any changes in the measurement instrument or changes in observers who are recording behavior threaten the internal validity of the experiment. In our example, there are no changes in the measurement instrument, and we are not relying on investigator observation, so this would not be a concern.

Statistical Regression: Sometimes it is appropriate to conduct experiments on subjects who start out with extreme scores on the dependent variables. In cases such as these, repeated measures may find those extreme scores moving toward the mean, even without exposure to the experimental variable. In our example, we are not attempting to recruit subjects at the extremes (those who have very strong or very weak opinions). While we might very well encounter subjects who fall into those categories, this should not be of concern, as we have randomly assigned subjects to one of the three groups. We have no reason to believe that, for example, all subjects with very strong opinions on a Supreme Court decision are in one group.

Selection Bias: Randomization eliminates the potential for systematic bias, but subjects may be chosen in other ways that threaten validity. Volunteers are often solicited for experiments conducted on college campuses. Students who volunteer for an experiment may not be typical of students as a whole. Volunteers may be more interested in the subject of the experiment and more likely to respond to a stimulus, or, if experimental subjects are paid, students in greater financial need may participate and may not be representative of other students. In our example, this could be a concern, depending on our recruitment techniques. If we are offering payment in return for participation, we would have to be aware of how that might affect our results. If the recruitment process were to identify the topic of the research, we might find that students concerned with the Supreme Court might be more inclined to participate than students who are not interested in the Supreme Court, once again possibly affecting our results.

Experimental Mortality: Subjects often drop out of an experiment before it is completed, and that can affect statistical comparison and conclusion. In our example, this would not be a concern, as we are conducting the entire experiment (pre-test/treatment/post-test) at one time. If we were to conduct the treatment and/or post-test at a different time, this could be a concern.

Causal Time Order: If there is ambiguity about the time order of the experimental stimulus and the dependent variable, the conclusion that the stimulus caused the dependent variable can be challenged with the explanation that the "dependent" variable actually caused changes in the stimulus. In our example, there is no ambiguity about the time order. The treatment in our experiment is structured in a manner to ensure that a change in the dependent variable can be attributed to the treatment.

Diffusion or Imitation of Treatments: In the event that experimental and control-group subjects are in communication with each other, it is possible that experimental subjects will pass on some elements of the experimental stimulus to the control group. In our example, if we were to conduct all the experiments at the same time, this would not be a problem. If we were to conduct some experiments at one time and further experiments at a later time, this might become a concern.

Controlling for Threats to External Validity

External validity refers to the generalizability of the results of your experiment. To what populations, settings, treatment variables, and measurement variables can the observed effect be generalized? The three major threats to the external validity of your experiment are people, places, or times. Are the people used in your experiment representative of the population you are interested in explaining? Did the location of your experiment prevent the results from being generalized to the population you are interested in explaining? Did the time at which you completed your experiment influence your results? For example, you might want to conduct an experiment in which you manipulated participants' feelings of security. You planned on conducting your experiment on September 14, 2001, 3 days after 9/11. The timing of your experiment would threaten the external validity of your results. Would you have had different results if your experiment had been conducted on September 10, 2001?

The external validity of your experiment can be improved in a number of ways. First, random selection of subjects can alleviate concerns about the generalizability of the results to people. Second, you can clearly describe the similarities between the subjects in your experiment and the population you are interested in explaining. Finally, you can replicate your experiment in a variety of settings, times and using a variety of subjects.

In our example, people are not a potential external validity problem. We are using college students as our subject,s and we want to be able to generalize our results to the general population. As college students are a part of the general population, this is not a concern; they are a representative sample of the general population. The location of our experiment (a college campus) would also not pose an external validity problem. Time should also not be a potential external validity problem unless an unforeseen event occurs. If, for example, a recent Supreme Court decision we planned on using in our experiment garnered a tremendous amount of media coverage right before our experiment, we might need to be concerned with the effects of this additional media coverage.

Controlling for Threats to Construct Validity

Construct validity refers to the degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based."[9] Construct validity is concerned with generalizing from our observations in an experiment to actual causal processes in the real world. Part of construct validity involves how completely an empirical measure can represent a construct or how well we can generalize from a measure to a construct. Threats to construct validity are problematic when researchers do not clearly specify what construct is to be represented by particular measures or experimental treatments. As Farrington, Ohlin and Wilson[10] note, "most treatments in existing experiments are not based on a well-developed theory but on a vague idea about what might influence [the dependent variable]…"

In our example, we have clearly defined our hypothesis and linked our empirical measures and treatments to our concepts. While we have not spent time discussing the connection to existing theory, that should be clearly stated in a research design. Clearly connecting theory, concepts, and measurements can eliminate or drastically reduce concerns about construct validity.

REQUIREMENTS

DATA

Very Low (1)

The data used for experimental research is generated by the researcher through the use of pre-tests, implementation of a treatment, and post-tests. That data, once collected, must be analyzed. Typically, statistical analysis is used to determine whether the treatment had a statistically-significant effect on the dependent variable, specifically using ANOVA procedures. Statistical analysis can also be used to describe the vast amount of data that was compiled by the researcher and to address *ad hoc* hypotheses that may have been developed by the researcher.

SET UP TIME

Average (3)

Time can be a serious consideration in experimentation. Considerable time is necessary to construct appropriate pre-tests and post-tests and to develop an appropriate treatment condition. Time also becomes an issue if there is a need for a time-lag between administering the pre-test and administering the post-test.

EXECUTION TIME

Short (0.25)

Statistical analysis of the data collected from the experiment does not take much time.

SKILL SET/EXPERTISE

Specialized-average (1)

[9] Research Methods Knowledge Base http://www.socialresearchmethods.net/kb/constval.php

[10] Farrington, D., L. Ohlin and J.Q. Wilson. 1986. *Understanding and Controlling Crime: Toward a New Research Strategy.* New York: Springer-Verlag.

While constructing pre-tests and post-tests does not require any specific skills, such as those required in statistical analysis, a thorough understanding of theory application, hypothesis development, and operationalization of concepts is necessary to ensure the development of an experimental model that addresses questions of internal, external, and construct validity. Failure to take these considerations into account results in unusable data. Familiarity with experimental design is required.

TOOLS
Minimal (0)
It is possible to use a software platform to run more sophisticated experiments that include audio visual components but for many research designs this is not necessary. Many experiments can be run using nothing more complex that paper and pencil. Analyses of results requires statistical computer programs such as SPSS, SAS or STATA.

COST
$ (5.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Experiments are particularly well-suited to questions that are process-oriented, as they enable control and observation of factors that are difficult to observe in the real world. For example, research questions that explore aspects of decision making, such as information processing and information form (text, audio, video, etc.), can be much more direct in an experimental setting, as the experimenter can observe (unobtrusively) the pattern of information acquisition, rather than relying on an individual's recollection and perception of how they reached a decision.

Experiments can also be very effective when the researcher is faced with a complex phenomenon for which little real-world data exists. This situation limits the ability to employ statistical models, as there are simply too many variables that need to be accounted for with the given data. In an experimental design, however, variables that may influence results but not be of central interest to the researcher can be held constant, while the effects of the variables of interest are examined.

FURTHER RESOURCES

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. New York: Houghton Mifflin Company.

Druckman. J., Green, D., Kuklinski, J., & Lupia, A. (2006). The growth and development of experimental research in political science." *The American Political Science Review*. 100(4): 627-35.

Gravetter, F., & Forzano, L. (2009). *Research methods for the behavioral sciences.* New York: Wadsworth, Cengage Learning.

Research Methods Knowledge Base (2011). *Construct validity*. Retrieved from http://www.socialresearchmethods.net/kb/constval.php

Research Methods Knowledge Base (2011). *External validity*. *Research Methods Knowledge Base.* http://www.socialresearchmethods.net/kb/external.php

Research Methods Knowledge Base (2011). *Internal validity*. *Research Methods Knowledge Base.* http://www.socialresearchmethods.net/kb/intval.php

GAME THEORY

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ❑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ❑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ❑ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Qualitative**
- ☑ **Quantitative**
- ❑ **Quantitative and qualitative**

Game theory is a formal branch of mathematics developed to deal with conflict of interest situations in social science. Since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* in 1944, the field has expanded enormously, and its impact has been particularly evident in the social sciences. Game theory can model economic situations at a variety of scales, from markets, where the choices of many affect the prices for all, to bargaining between individual buyers and sellers. Political situations, both domestic (elections, legislative voting in committees) and international (bargaining, the decision to go to war), can all be modeled using game theory, as have a broad range of group interactions across a diverse range of settings.

METHODOLOGICAL OVERVIEW

WHAT IS GAME THEORY[11]

"Games" is a scientific metaphor for strategic situations, which are a subgroup of social situations. **Social situations** involve interactions between individuals, and to understand social situations, we need a theory that can explain how individuals' decisions are interrelated and how those decisions combine to generate outcomes. This is what game theory does.

A **strategic situation** is a setting where the outcome depends on the actions of all players who have opposed or, at best, mixed motives. Game theory utilizes a highly structured series of moves to

[11] The structure of the discussion in this section is taken from Morrow, J. D. 1994. *Game Theory for Political Scientists* Princeton NJ: Princeton University Press.

show how an individual player's utility depends not only on his choices, but on the choices made by other players. The choices players make are also influenced by the social framework, or structure in which they are interacting. **Social structure** includes such things as capabilities, resources, and individuals' perceptions of their options. Game theory provides a way to formalize the relevant elements of a social structure and examine how they affect individual decision making. Put more simply, a game is a situation in which there are two or more players, each of whom responds to what the other does, or what each thinks the other might do.

WHY USE GAME THEORY?

The primary advantage of formal modeling approaches, such as game theory, is the rigor and precision of argument they require. The social world is immensely complex, and many verbal theories and arguments to explain social phenomenon fail to specify their assumptions and assertions. Thus, the conclusions drawn from those arguments will not hold for all cases (will not be generalizable), because they depend on a set of unstated assumptions. Writing down an argument or theory formally forces the modeler to articulate precisely what the assumptions of that model are.

This clarity of structure allows the modeler to identify all conclusions that follow from a model's assumptions, even those not initially considered. These unexpected conclusions may suggest new avenues of enquiry that lead to conclusions beyond the initial intuition underlying the model. Game theoretic models can also produce conclusions that are at odds with our observations of the real world. The logical structure of a model allows the modeler to add to or modify a model in order to derive new conclusions that can explain these observations.

The rigor and precision that game theory brings to a topic, however, requires sacrificing the detail that often lends verbal arguments their realism. Game theory and all formal models strive to pare down social phenomenon to their simplest components without losing their essential nature. Any description of the social world, no matter how dense and descriptive, is a simplification of an endlessly complex reality. How we choose what we include and what we exclude whenever we describe or explain a phenomenon is a theoretical statement; we account for only those things we consider to be relevant. Modeling requires us to be explicit about every one of these choices and assumptions, and the need for simplicity creates greater generalizability and maintains a focus on the basic mechanisms behind a phenomenon of interest. Political scientist James Rogers' analogy between models and maps explains how these abstractions are useful when we try to understand the social world.

> …the very unrealism of a model, if properly constructed, is what makes it useful. …If one compares a map of a city to the real topography of that city, it is certain that what is represented is a highly unrealistic portrayal of what the city really looks like. The map utterly distorts what is *really* there and leaves out numerous details about what a particular area looks like. But it is precisely *because* the map distorts reality – because it abstracts away from the host of details about what is really there – that it is a useful tool. A map that

attempted to portray the full details of a particular area would be too cluttered to be useful in finding a particular location or would be too large to be conveniently stored. [12]

RATIONAL CHOICE AND MODELING[13]

A crucial assumption that underlies game theory is that all players in a game are rational. That is, "they have goals and they attempt to realize those goals through their actions" (Morrow 1994: 17). Morrow delineates the rational choice approach as follows:

- People have goals they seek to achieve, but the morality or objective value of those goals is not judged.
- Although the structure of the game does constrain the choices available to players, it is assumed that there is some freedom of choice.
- Individuals choose actions they believe will help them achieve their goals.
- Reality is deliberately simplified and abstracted in order to help clarify complex interactions.

HOW DO YOU CREATE A GAME?[14]

There are several central components that create a game; the choices available to players, how those choices lead to outcome,s and how the players evaluate those outcomes. The choice of specific game structure is determined by the social theory (or set of assumptions) we want to apply. Game theory cannot tell us whether a theory is an accurate description of the world; rather, it lays out the behavior we should expect from players as a consequence of that theory (or set of assumptions). As most social situations involve at least two players, **two-person games** are the simplest interactive situations of general concern to social scientists[15]. Games that involve more than two players are referred to as **n-person games**.

DEFINING A GAME[16]

Games of strategy arise in many different social contexts and, accordingly, have a variety of different salient features that must be captured in the game structure. These features can be

[12] Rogers, James. 2006. "Judicial Review Standards in Unicameral Legislative Systems: A Positive Theoretical and Historical Analysis." *Creighton Law Review* 33(1) 65-120

[13] The following discussion of rationality draws on Morrow 1994

[14] The discussion in this section is taken from Zagare, F. 1986 "Recent Advances in Game Theory and Political Science" in Samuel Long (ed.) Annual Review of Political Science Norwood NJ: Ablex Publishing Corporation

[15] The simplest type of game is a one person game (also called a game against nature), where a single player must make a decision in the context of an environment that is either indifferent or neutral, but this lacks the strategic component crucial to games that model social situations.

[16] The organization of this section is adapted from Dixit, A. and S. Skeath 1999. *Games of Strategy* New York: WW Norton & Company

grouped into a few defining categories that enable us to identify a set of pure form games. The following questions can be used to help classify a particular gam, or to choose an appropriate game for a specific research question.

1. Are the moves in the game sequential or simultaneous?

   This distinction is important because the structure of play affects the type of interactive thinking required. In **sequential games,** players take turns to move, and so their moves are determined by calculation of future consequences. In **simultaneous games,** all players move at the same time, so the task is to determine what the opposing player is going to do immediately. Sequential games also tell us when it is an advantage to move first and when second.

2. Are the interests of players in complete conflict, or are there some common interests?

   Games can be distinguished according to the extent to which the interests of the players diverge. Games in which conflict is total and players' interests are diametrically opposed are known as **zero-sum**, or **constant-sum games**. Those games in which players have both competing and complementary interests are known as **nonzero-sum games**. Economic and social games are usually not zero-sum, neither are wars or strikes. The Pyrrhic victory that characterizes total nuclear war is the most striking example of a nonzero-sum situation; one in which there is only losers.

3. Is the game played once or repeatedly, and do the players change or remain the same?

   If a game is played once (**one-shot game**), players have less information about each other, and their strategies are short term by nature. In **iterated games,** players develop ongoing relationships and reputations, and such future considerations can change the type of strategy that optimizes results for players. Consequently, a game that is zero-sum in the short run may have scope for mutual benefit in the long run.

4. Do the players have full and equal information?

   Control of information (what information you release and when) can provide players with significant strategic advantage. Players attempt to reveal **good information** (information that draws advantageous responses from other players) and conceal **bad information** (information that is to the player's disadvantage). The lack of **complete information** that characterizes most games creates incentives for players to exaggerate or lie about their capabilities or intentions to gain a strategic advantage. For example, during arms control negotiations, states have a strong incentive to underreport existing weapons counts in order to gain an advantage at the end of the bargaining.

5. Are agreements to cooperate enforceable?

Finally, the rules assumed to govern play can also be used to distinguish between games; specifically, games in which binding agreements are precluded (**non-cooperative games**) and those where binding agreements are possible (**cooperative games**).

It is this last question that is most important from a theoretical perspective, and separate theories have evolved for cooperative and non-cooperative games.

Cooperative games allow for enforceable joint-action agreements. Agreements to cooperate can succeed when there is immediate implementation and oversight, but this situation is rare. Players most generally only cooperate with an agreement that is not in their individual interest when their actions are observable by others and there is a third party capable of enforcing compliance; for example, a court ruling within a sovereign state.

Non-cooperative games start from the assumption that players are unable to communicate or negotiate binding agreements with one another. In these games, cooperation will emerge only when it is all players' separate and individual interest to continue to take the prescribed action. This makes this branch of game theory particularly well suited for analyzing both situations where there are institutional mechanisms that block direct communication between players (anti-trust laws) and those where institutional enforcement mechanisms do not exist (the international system).

ELEMENTS OF A GAME

Specifying the structure of a game involves identifying all the strategies available to all the players, their information, and their objectives and preferences.

Players: A player can be either an individual or a group of individuals functioning as a single decision-making unit. Individuals or groups become players in a game when their decisions, combined with the decisions of at least one other player, produce an outcome.

Strategies: A strategy for a player is a full set of instructions as to how to play in the game. A **pure strategy** is a complete contingency plan that specifies a choice for the player in every possible situation. A **mixed strategy** involves the use of a particular probability distribution to select one pure strategy from among a subset of a player's pure strategies. **Dominant strategies** are unconditionally best strategies that give a player a better outcome, regardless of the strategy chosen by the other player.

Utilities: Players are assumed to be able to evaluate and compare the consequences associated with each outcome and assign a numeric value, a utility, to each outcome. This utility can be thought of as the **payoff** associated with that outcome for the player. If the outcomes of the game are probabilistic in nature, then the payoffs are a measure of **expected utility**, the value placed on the outcome weighted by its probability.

Preference: The utility a player assigns to each outcome in a game creates a preference ordering among those outcomes. When the values reflect only a rank ordering of the outcomes, they are referred to as **ordinal utilities**; where values denote both order and intensity of preference, they are known as **cardinal utilities**.

Outcomes: For every game, there is a set of possible outcomes. These outcomes represent all possible end points for the strategic interaction that the game is designed to capture. Each outcome is associated with a set of payoffs that indicates the utility of that outcome for each player.

Representing the Structure of a Game

Game theorists have developed two basic structures for abstracting and representing the essential features of an interactive situation – the game. Specifying the structure of a game involves identifying all the strategies available to all the players, their information, and their objectives and preferences. There are two ways to visualize games: normal and extensive form.

Normal form: Games can be represented by a simple payoff matrix (Figure 1) that shows the players, strategies, and payoffs (utilities) associated with each possible outcome. Games in which both players move simultaneously or without knowledge of any previous moves by the other player can be illustrated in a normal form.

Prisoner's Dilemma: Figure 1 is a normal form representation simultaneous, non-cooperative game, Prisoner's Dilemma (PD). In this canonical game, two accomplices have been arrested, but the police have insufficient evidence for a conviction and must rely on a confession. The police put each prisoner in an isolation cell and give both the same offer: Confess or remain silent. If they confess (defect) and agree to testify against their accomplice, and their accomplice remains silent (cooperate), the defector goes free and the other receives a six-year sentence. If both remain silent (cooperate), each are sentenced to one year on a minor charge. If each implicates the other (defects), both receive a three-year sentence. In the classic one-shot version of this game, the dominant strategy is defect, and the only possible equilibrium for the game is for both players to defect. However, this is also sub-optimal strategy, as both would have obtained a better outcome if they remained silent.

		Prisoner Y	
		Cooperate (Stay silent)	Defect (confess)
Prisoner X	Cooperate (Stay silent)	1, 1	6, 0
	Defect (confess)	0, 6	3, 3

Payoffs for row player (Prisoner X) listed first.

Figure 1: Prisoner's Dilemma

PD is a **canonical game** because the conflict between individual and collective interest that the game neatly encapsulates is the same conflict that lies at the heart of many important social situations, and so it has implications for political and other types of social systems. For example, a decision on whether or not to agree to arms reduction or non-proliferation treaties, to join an international institution, such as the UN or WTO, are all PD-type choices that pit an individual's (or individual state's) interests against the interests of a wider community. According to Kuhn[17],

[17] Kuhn, Steven, "Prisoner's Dilemma", *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2009/entries/prisoner-dilemma/>.

> The "dilemma" faced by the prisoners here is that, whatever the other does, each is better off confessing than remaining silent; however, the outcome generated when both confess is worse for each than the outcome they would have obtained had both remained silent. A common view is that the puzzle illustrates a conflict between individual and group rationality. A group whose members pursue rational self-interest may all end up worse off than a group whose members act contrary to rational self-interest.

Extensive form: **Game trees** can be used to formalize non-cooperative games in which moves are made sequentially. Nodes within the tree represent a point where a player gets to move (decision node), and terminal nodes represent end points (outcomes) of the game. Each terminal node is associated with a payoff for each player that indicates the utility for each player of the game ending at that particular node.

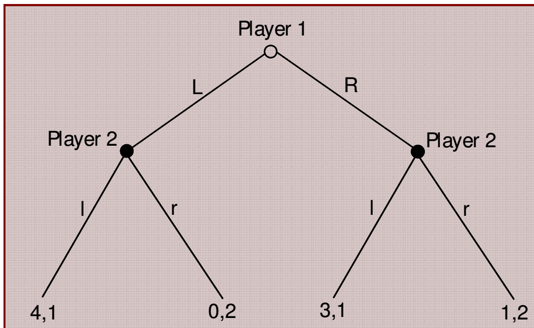


Figure 2: Two player extensive form game

The game shown in Figure 1 is a two-player game in which each player gets one move, starting with Player 1. In this game, Player 1 can chose either left (L) or right (R), after which Player 2 can chose either left (l) or right (r).

The strategies are slightly more complicated than those for a normal form game, as Player 2's strategies are contingent on the choice made by Player 1.

- Player 1 strategies: {L,R}
- Player 2 strategies: {(l if L, l if R),(l if L, r if R),(r if L, l if R),(r if L, r if R)}

PREDICTING OUTCOMES: THE IDEA OF EQUILIBRIUM

The idea of equilibrium follows from the rational choice approach to understanding social phenomena. An equilibrium outcome is an outcome in which no player has an incentive to unilaterally switch strategies to reach a different outcome. Behavior at equilibrium is stable because no player, given their current position and knowledge, can unilaterally improve their situation. It is important to distinguish between equilibrium outcomes and **optimal outcomes**. Equilibria are not fair or desirable outcomes by any ethical criteria; in fact, they are often both greatly unfair to one player or reflect socially suboptimal outcomes. A designation of equilibrium is simply the identification of a point in the model at which no player wishes to change the choices it has made.

Equilibria serve as the predictions of a model. Identifying these regularly occurring outcomes, therefore, is a precondition to the determination and specification of general principles of social behavior. In classical non-cooperative game theory, Nash's equilibrium concept is standard. **Nash equilibrium** occurs when no player can benefit immediately from a unilateral strategy switch.

REQUIREMENTS

Game theory tests expectations and finds solutions to strategic situations that are abstract in nature. Unlike other methodologies, such as statistical and case study, there it provides no indication of the empirical usefulness of the game's solutions; that is, the extent to which actors in the real world act in accordance with the expectations and assumptions of the game. Much game theoretical work, particularly microeconomics, involves the use of incentivized laboratory experiments to test game theoretic models against real-world behavior. In political science, case studies have more commonly been used as real-world support for a game's solutions. However, in the area of foreign policy decision making and voter behavior, the advantages of experimental, over case study, validation are increasingly recognized. Much of this work focuses on how individuals violate the assumption of rationality, creating outcomes that are not predicted by the game theoretic model. Testing the expectations of a game theoretic model requires additional resources, time, and data.

DATA

Very Low (1)

One of the advantages of game theory is that the data requirements are fairly small compared to other techniques. The analyst needs to know enough about a strategic situation of interest and the players involved to determine the strategies available to players, the outcome set, and enough about the players' preferences to assign payoffs to those outcomes. As discussed above, if there is a desire to test the solutions derived from a game, then data generated from either computer simulations or experimental tests will be required.

SET UP TIME

Short (2)

As game theory analyses do not require any data collection effort, the majority of time required is related to determining the appropriate structure and payoffs for the game and identifying the equilibrium outcomes and their implications.

EXECUTION TIME

Short (0.25)

As game theory analyses do not require any data collection effort, the majority of time required is related to determining the appropriate structure and payoffs for the game and identifying the equilibrium outcomes and their implications. However, if the game is to be tested using experimental data, then time requirements will increase considerably.

SKILL SET/EXPERTISE

Specialized-advanced (2)

Game theory does require the analyst to be trained in the fundamentals and application of the technique. While games, such as the Prisoners' Dilemma discussed above, are, in and of themselves, simple games, this simplicity belies the level of experience and understanding required to accurately undertake a game theory analysis. Some background in mathematics is also required for setting up and solving more complex game structures. In addition, if there is a desire to compare the game solutions to real world data, then it will also be necessary to involve someone with experience in the design and conduct of social science experiments.

The primary resource required to construct a game theory model is an experienced game theorist who can design a game that accurately captures your research question and the parameters of the game.

The strength and usefulness of the solutions generated from a particular game are dependent on the accuracy of the games and understanding of the preferences of the players. Determining preferences and understanding how a particular player (for example, a country or a specific political leader) would interpret and respond to a specific strategic situation is, therefore, critical. This is one area where SME's can be of great use to the game designer.

TOOLS
Minimal (0)
Game theory requires no specific tool or programs. However, if game outcomes are to be tested against data some statistical software will be necessary.

COST
$ (0.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Among the issues discussed in game theory are:

1. What does it mean to choose strategies "rationally" when outcomes depend on the strategies chosen by others and when information is incomplete?

2. In "games" that allow mutual gain (or mutual loss), is it "rational" to cooperate to realize the mutual gain (or avoid the mutual loss) or is it "rational" to act aggressively in seeking individual gain regardless of mutual gain or loss?

3. If the answers to (2) are "sometimes," in what circumstances is aggression rational, and in what circumstances is cooperation rational?

4. In particular, do ongoing relationships differ from one-off encounters in this connection?

5. Can moral rules of cooperation emerge spontaneously from the interactions of rational egoists?

6. How does real human behavior correspond to "rational" behavior in these cases?

7. If it differs, in what direction? Are people more cooperative than would be "rational?" More aggressive? Both?

Non-cooperative game theory has been applied to the study of two-party and multi-party elections, legislative decision-making, bureaucratic politics, international crises, deterrence, and international cooperation. More generally, questions of how political institutions work, why they exist and change have been addressed game theoretically, as has communication in a variety of settings.

FURTHER RESOURCES

Dixit, A. and S. Skeath 1999. *Games of Strategy* New York: WW Norton & Company

Morrow, J. D. 1994. *Game Theory for Political Scientists* Princeton NJ: Princeton University Press.

Zagare, F. 1986 "Recent Advances in Game Theory and Political Science" in Samuel Long (ed.) Annual Review of Political Science Norwood NJ: Ablex Publishing Corporation

Polak, Ben "Game Theory" http://oyc.yale.edu/economics/game-theory/ This course, available online, is an introduction to game theory and strategic thinking.

NARRATIVE ANALYSIS

Identify the levels, data observations and forms for which this method is most suitable.

Level

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Time-series** : multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ☑ **Qualitative**
- ❑ **Quantitative and qualitative**

INTRODUCTION

Typically associated with literature and story-telling, narratives and the analysis thereof provide a powerful, qualitative tool to explore different facets of a population's belief system. As an applied technique, narrative analysis has been more recently used across multiple fields, including sociology, psychology, and health sciences (Hyvärinen, 2006). Narrative analysis makes use of the fact that a narrative is affected by and reflects society and culture and how people in those societies view the world. This can allow a skilled researcher to examine questions that other qualitative and quantitative methods cannot accurately evaluate with narratives filling gaps as a rich source of information about perspective, experiences, and events.

METHODOLOGICAL OVERVIEW

WHAT IS A NARRATIVE?

A narrative goes beyond what people simply call a 'story'. Roland Barthes (1975) wrote, "narrative is present in myth, legend, fables, tales, short stories, epics, histories, tragedy, *drame* [suspense drama], comedy, pantomime, painting (in Santa Ursula by Carpaccio, for instance), stained-glass windows, movies, local news, conversation" (p. 237). Most definitions of narrative note that a narrative covers a sequence of connected events. So while a narrative does not have to be a particular length, it does require the connection of events that cover a beginning, middle, and end.

Narrative analysis does not have a formalized or specific method. Linguists use narrative analysis to explore vernacular language, sociolinguists use narratives to study the stories of relationships and experiences from the micro-level (personal) to the macro-level (society), and media analysis explores the narratives of film or other media. Structural narrative analysis is a means to explore narratives in different genres, how a narrative reflects culture (or vice versa), and the relationship between a narrative and its context. This structural approach to narrative analysis has been used to study texts from Russian folk tales (e.g., Propp, 1968) to Taliban night letters (e.g., Johnson, 2007).

Advantages of Narrative Analysis

Storytelling is pervasive in human culture as stories about experiences, perspectives, thoughts, feelings, and opinions are routinely shared. Our stories or narratives are a way to convey information that represents more than a simple recitation of a timeline of events. Narrative analysis is a means to analyze the very human act of storytelling. To some scholars, narrative analysis may be the only way to preserve information that other methods leave behind (Smith, 2000). The language used to tell a story is important because language reveals our "thoughts, feelings, and sensory experiences into a shared symbolic form." Additionally, narrative analysis can be applied to the stories, writings, diaries, etc. of people who are no longer living. The language used within a narrative can also expose more information than the author may have intended (e.g., insights into opinions, emotions, etc.) (Smith, 2000). Therefore, narratives reveal more information than what could be discerned in self-report questionnaires or observational studies.

Limitations of Narrative Analysis

Narrative analysis is not easily definable. In some scholarly fields, narratives are defined as verbal recitations of personal experiences or interviews. Elsewhere, narratives are interpreted through the lens of literature such as dramas, poems, or cinematic scenes. Due to the variety of narrative interpretation, the methods of analyses vary widely. Additionally, social science researchers also define a narrative as a continuing and pervasive underlying story (or stories) within a culture. With each definition of narrative and narrative analysis, the actual study and method varies. This grey definition and understanding of narrative analysis can be confusing and lead to difficulty in creating a reliable (or reproducible) method.

In addition to the vague definition and procedure of this qualitative method, narrative analysis is a time-consuming process. Depending on availability of source materials and the ability to collect the appropriate narratives, even the data gathering process will require time. Sufficient time will also need to be allocated for reading and data collection and analysis from the narratives. This amount of time will vary according to the number of documents collected and the length of the text of each document.

Finally, narrative analysis, as a qualitative method, is generally not considered to be replicable (ability for multiple researchers to do the same study and receive similar results). Different analysts may not collect the same data from narratives, identify the same themes, or come to the same conclusions.

LITERARY FOUNDATIONS OF NARRATIVE ANALYSIS

At its foundation, narrative analysis utilizes theories from the literary field. Freytag's pyramid is a common literary tool applied to the analysis of stories, often taught in secondary school English or literature classes. Freytag's pyramid is used to define what is or is not a story and thus is important to identify a narrative. Later literary scholars added to the definition of narrative by defining the difference between the story and how that story is told.

Freytag's Pyramid

Freytag's pyramid (Figure 2) is often used to describe the structure or elements that make up a story; therefore, it is often used to describe the evolution of a narrative. The pyramid is made up of five parts: exposition, rising action, climax, falling action, and denouement. The **exposition** is the introduction of the main characters, situation, and basic conflict. **Rising action** is the part of the story where the main character or protagonist becomes aware of the conflict and encounters complications. The **climax** is the turning point or period of change in the story. **Falling action** describes the action or conflict resolution. Finally, the **denouement** marks the end of the story where the main character learns something from the conflict and has somehow changed from the beginning of the story (Freytag, 1900).

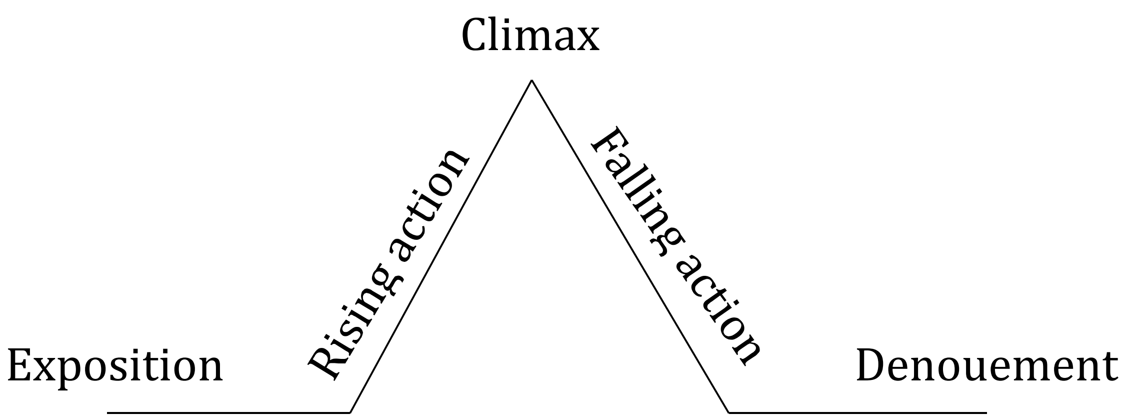


Figure 2. Freytag's pyramid.

Narrative as a Story and a Plot

A narrative is basically made up of two parts: the series of events (plot) and how that information is conveyed to an audience (story). Literary scholars of the Russian formalist school in the early 20th century emphasized the importance of the audience's participation in the construction of a narrative. They added two terms: *syuzhet* (or *sjuñet*), which represents the series of events that explicitly happen in the narrative text, and *fabula,* which represents how the audience constructs or fills in the gaps between these events. Later, literary scholars of the French structuralist school divided the plot into the series of events (*discourse*) and then the act of narrating. Franzosi (1998) represented the overlapping terms that make up a narrative (Figure 3). A narrative must have this series of events or succession of events and they must be linked by some kind of change in the action (Todorov, 1990 as cited in Franzosi, 1998). Later, literary scholars added to the idea that a narrative is made up of the plot of the story and how the story is told, referring to the narrator and the language used to tell the story.

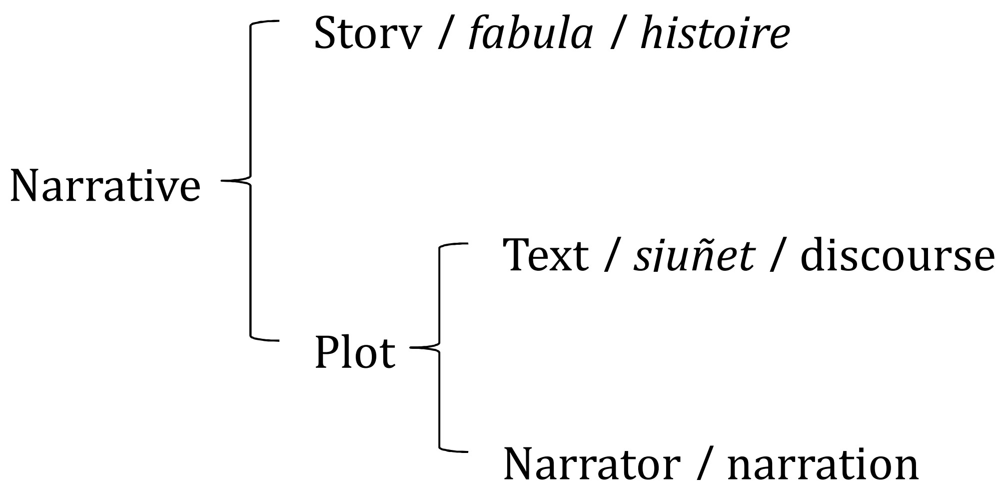


Figure 3. Breakdown of a narrative. Adapted from Franzosi, R. (1998, p. 520).

William Labov and Joshua Waletzky in the late 1960s brought a sociolinguistic dimension to narrative analysis with their identification of the five elements of a narrative and how language is used within these elements. While Labov and Waletzky are not literary scholars, their work on the structure of narrative is an important concept to the modern application of narrative analysis.

NARRATIVE STRUCTURE

Labov and Waletzky (1966) described their analytical framework as a formal and functional analysis of narrative. Formal[18] in that analysis is "based upon recurrent patterns characteristic of narrative from the clause level to the complete simple narrative" (p. 12). Functional in that analysis captures the sum of the experience and that the narrative relays personal experience and opinion within the social context.

Additionally, they identified five elements of a narrative: orientation, complicating action, evaluation, resolution, and coda. Narratives generally have some kind of **orientation** that clues the listener/reader into the basic information of the story such as person, place, time, and behavioral situation. The **complicating action** is generally the bulk of the narrative in which an event or series

[18] See Labov & Waletzky (1966) for examples and further explanation on how they broke down sentences into clauses for their analysis.

of events occur. The **evaluation** within a narrative refers to how the narrator talks about the events in the story, the overall attitude of the narrator, and how the narrator refers to events that have not yet happened or events that could have happened. The **resolution** is the end of the action and the evaluation. Finally, some narratives end with a **coda**, an element contained within one sentence, in which the narrator brings the reader to the present (e.g., happily ever after) (Labov & Waletzky, 1966, p. 40).

OTHER IMPORTANT CONCEPTS

Smith (2000) reviews two main concepts that are important within a narrative. **Perspective** is the idea that narratives are told from a particular point of view and also an opinion of what is significant or worth being repeated to the audience. **Context** refers to the influences on the narrator, the ways in which the narrator creates the narrative, and how the final texts turn out (Smith, 2000, p. 328). There is an important distinction between Smith's use of **context** and the use of "context"[19] to refer to the state of what is happening outside of the narrative. *The context within the narrative, as well as the context in which the narrative occurs, are both important to an analysis.*

Intertextuality builds on the idea that the context in which the narrative occurs is critical to understanding the narrative. **Intertextuality** is defined by Fairclough (2003) as "how texts draw upon, incorporate, recontextualize and dialogue with other texts." Intertextuality refers to the idea that a text includes more than just the words within. In other words, an author may implicitly or explicitly refer to outside knowledge (or other texts) that he or she may not introduce or discuss in the text. Often these implicit or explicit references are cultural or contextual. For example, a news story from the United Kingdom may make a reference to the Troubles, without giving any further explanation. An American reader born in the mid-1980's may not know that the Troubles refer to a roughly 30-year period of conflict and violence in Ireland. In addition to not understanding the reference, the American reader may not even realize the implications of not understanding the background in relation to the text he or she is reading. When an analyst is unfamiliar with the cultural, social, or political context of the narrative, he or she may overlook or completely misunderstand references within the narrative.

The concepts of perspective and context, as well as intertextuality, are important to narrative analysis and add a richer understanding of the narrative and its insight into opinions, sentiment, events, and even society in some respect. A narrative does not exist in a vacuum, and examining context and intertextuality get at the bigger picture. The internal structure of the narrative must also be examined and a method to do so is outlined in Barthes (1975).

[19] Merriam-Webster defines context as the interrelated conditions in which something exists or occurs.

NARRATIVE METHOD

Similar to content analysis, there is no one exact or accepted method of analyzing narratives. Multiple scholars have proposed different methods and typologies (see Riessman, 2005). The qualitative nature of narrative analysis makes a methodology difficult to define. While this can be beneficial, allowing for flexibility and adaptation, it also means that no two analyses are alike. Therefore a procedure and the interpretation are not reproducible. With this in mind, the following narrative method is summarized from Roland Barthes' (1975) structural narrative analysis method. This method lists the possible functions available for analysis in a narrative. Generally, the functions or elements of a narrative worth examination are the story, scene, events, characters and their actions, and motivations.

Defining the Functions

Barthes (1975) notes that a narrative is made up of functions and that every function is meaningful and plays a part in the narrative. First, a narrative must be broken down into the appropriate units. A researcher must decide what the smallest narrative unit should be for analysis. The narrative unit is defined based not only on its meaning but also on the function the unit of meaning plays in the story.

Next, these narrative units can be organized into a small number of formal classes. Two broad classes are distributional and integrative units.[20] **Distributional** units are the action events that occur within a text. Distributional or functional units can be broken down into nuclei and catalyses functions. **Nuclei** functions are the actions that either start or resolve another action and the **catalyses** are the mundane actions that occur between nuclei. For example, the overall action was that Robert opened the door. The nuclei would therefore be the action of opening the door and the catalyses would be the smaller actions that took place but were not stated, e.g., walking to the door, turning the doorknob, etc. **Integrative** units are the parts of the text that refer to diffuse concepts like personality or atmosphere. Integrative units can be either indices or informants. **Indices** are implicit information that a reader must infer or deduce from the text (e.g., 'lapping of water' probably means a body of water is nearby); while **informants** are explicit bits of information that are stated (e.g., a person's age).

This review of integrative and distributional units covers only the surface of these concepts. A review of Barthes' (1975) article would be beneficial for a more in depth look at these theoretical concepts. A research project that strictly adheres to the structural method would analyze the functions of a narrative at such a specific level. Other narrative analyses review the more general elements involved, such as character, action, story, motivation, etc.

Defining the Actions

[20] See Barthes, 1975 for more in depth discussion.

The actions, events, or **characters** (*actants*) should be defined. Characters can be defined personally by name or by their role. For example, Propp identified common roles such as villain, hero, false hero, or helper. Defining the roles of the character opens up analysis of motivation or reason why characters act the way they do. Other related active elements in the narrative to be studied are the actions of the characters and the events that occur within the narrative.

Defining the Narrative

The narrative exists as an object in and of itself. This means that both the actual narrative and the *act* of narrating a story are elements to be defined and analyzed. Barthes writes that the giver and recipient of the narrative need to be probed and examined. The narrator[21] can be described on the basis of how he/she/it is represented (first-person, omnipresent, etc.), tone, or what the narrator leaves out. Additionally, Barthes writes that the narrative situation should be examined. The **narrative situation** is the idea that a narrative makes sense in the context in which it is consumed.[22] A narrative should be examined from the time and place in which it was written. For example, an article describing the state-of-the-art in computers will be completely different in 1951 than 2011. Also, the reader or the intended audience's situation is an important context to examine. Questions, such as *who* is the narrative intended for and *how* and *why* is it consumed, should be asked.

REQUIREMENTS

DATA

Low (2)

Data will be at the discretion of the researcher based on both the research question and what narratives are available for analysis. Narratives are generally thought of in terms of documents of actual text. However, Barthes believes narrative analysis can be done on anything that tells a story, even a painting. Because narratives could be found in so many forms, there is no set or minimum length of time a narrative analysis should cover. Narrative analysis can often be longitudinal as texts are drawn from different points in time, but a time period need not be defined or finite.

SET UP TIME

Average (3)

The time required to conduct narrative analysis will vary from project to project and will depend on multiple factors. If texts are not readily available, acquiring materials to analyze can be a lengthy and time-consuming process.

EXECUTION TIME

Long (1)

[21] The narrator is not the same thing as the author.

[22] See discussion on intertextuality.

Analysis will depend on the kind and length of text used. Imagery or headline analysis may be faster than analysis of stories or news documents. Also to be considered is the number of texts to be analyzed. If analyzing ten documents as opposed to 100, then the overall time requirement for analysis can be adjusted appropriately.

SKILL SET/EXPERTISE
Specialized-average (1)
Narrative analysis requires a thorough literature review to understand the context of the narrative research question. There should be a research plan in place to guide study and prevent mindless qualitative analysis. A researcher must be prepared to search for appropriate and available texts for analysis and be able to make adjustments as necessary.

Narrative analysis does not rely on numbers or statistics to convey information and therefore the write-up is critical in communicating the results and conclusions. Solid writing skills will be essential to the process.

TOOLS
Minimal (0)

COST
$ (7)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

The 5D framework asks three general questions: 1) How effective is current US force posture for achieving policy objective?; 2) What would be the optimal force posture to achieve a specific policy objective?; and 3) What strategy is optimal to achieve the objective?

1. *How effective is current US force posture for achieving policy objective?*

   Our current force posture does not appear to draw on narrative analysis. For instance, in the Nuclear Terrorism Deterrence project several years ago, no narrative analysis was done.

   One effort that comes close to narrative analysis is The Islamic Imagery Project done at West Point (Combating Terrorism Center, United States Military Academy, 2006). It is an analysis of pictorial imagery used by Jihadists. The significance of particular images (e.g., animals, colors, people, and other symbols, like hands in prayer) are discussed within the context of Jihadi images, not just images associated with Islam or particular countries or groups.

   The most notable government funded narrative analysis come from Tom Johnson. Johnson (2007) examined Taliban 'night letters' to analyze the how the Taliban used information to gather support from local populations. He writes that some folklore reinforces ideas that act as a form of social control. Johnson draws on examples of these 'night letters' to identify particular themes within Taliban communication. Understanding these themes is important to support information and psychological operations, add to the understanding of the culture, and to inform how future mobilization or reconstruction should unfold (Johnson, 2007).

One noteworthy narrative analysis from academia is Post, Sprinak, and Denny's (2003) study of personal narratives of incarcerated terrorists. The research team used semi-structured interviews to examine the opinions, motivations, and reasoning of the terrorists.

2. *What would be the optimal force posture to achieve a specific policy objective?*

Narratives are fundamental to human communication and the transmission of ideas and culture (Bruner, 1991; Langellier, 1989; Smith, 2000). People betray both sentiment and intent through the rhetoric they use. Therefore, maintaining a robust national capability for the analysis of narrative, for dealing with adversaries, allies, states and non-states, is necessary. These capabilities have likely homes in multiple places: the Department of State, the non-military intelligence community, as well as the military intelligence, planning communities and civil affairs arenas.

Fielded units engaged in communication with adversaries and host populations or governments should have some basic familiarity with content analysis and narrative analysis. Units on the ground may not have in theatre time and resources for analytic study, however, awareness of the concepts and information available should be part of their operational planning. They may not have the resources and time in theatre to conduct such analyses, but they at least need to be aware of the importance of narrative and aware of what kinds of information they could gather and what kinds of analyses that could inform their own operations.

Large scale content analysis and in-depth narrative analyses requires time and technical resources. Reach back cells should be maintained in order to supply such analyses. One example is the Human Terrain System (HTS) regional reach back cells in the US, which support the research of fielded human terrain teams.

3. *What strategy is optimal to achieve the objective?*

The first step is assessment of current military and IC capabilities for content analysis and narrative analysis. Who does it now? Where? What are their challenges and needs? A gap analysis should identify who should have these capabilities and how their personnel could be trained or new personnel hired to fill those gaps.

Since content and narrative analysis are relatively new and developing fields, achieving a robust capability for content and narrative analysis in the military and the intelligence community also requires government funded R&D of the nation's brain trust in academia and industry. Research should support both theoretical advances in understanding human communication and behavior, as well as methodological advances in analysis of narratives.

Milestones for success in this research need to be defined, and as specific projects mature, clear pathways for transition to military training and operations must be defined. This would include the identification of military and non-military specialties where these capabilities are most needed and best used, and plans for how new methods, techniques, and procedures would be trained, accredited, and incorporated into military and intelligence community business.

FURTHER RESOURCES

Basic References

Barthes, R. (1975). An introduction to the structural analysis of narrative. (L. Duisit, Trans.). New Literary History, 6 (2), 237- 272.

Franzosi, R. (1998). Narrative analysis: Or why (and how) sociologists should be interested in narrative. *Annual Review of Sociology, 24*, 517-554.

Labov, W., & Waletzky, J. (1966). Narrative analysis: Oral versions of personal narrative. In J. Helm (Ed.), *Essays on the Verbal and Visual Arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society* (pp. 12-44).

For Further Reading

Casebeer, W. D. (2006, March). Military force and culture change: Systems, narrative, and the social transmission of behavior in counter-terrorism strategy. (Master's thesis, Naval Postgraduate School). Retrieved from http://www.nps.edu/ programs/ccs/docs/pubs/casebeer_thesis.pdf

Mishler, E. G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History, 5*(2), 87-123.

Upal, A. (Ed.). (2010, April). *The role of narratives in shaping people's social identity beliefs* [Technical Report 2010-062], Workshop proceedings from Defense R&D Canada-Toronto.

WORKS CITED

Barthes, R. (1975). An introduction to the structural analysis of narrative. (L. Duisit, Trans.). New Literary History, 6 (2), 237- 272.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*(1), 1-21.

Combating Terrorism Center, United States Military Academy. (2006, March). *The Islamic imagery project: Visual motifs in Jihadi internet propaganda*. Retrieved from http://www.ctc.usma.edu/wp-content/uploads/2010/06/Islamic-Imagery-Project.pdf

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. London: Routledge.

Franzosi, R. (1998). Narrative analysis: Or why (and how) sociologists should be interested in narrative. *Annual Review of Sociology, 24*, 517-554.

Freytag, G. (1900). *Technique of the drama : And exposition of dramatic composition and art* (E. J. McEwan, Trans.) [Internet Library version]. Chicago, IL: Scott, Foresman, and Co. Retrieved from http://www.archive.org/stream/freytagstechniqu00freyuoft# page/n3/mode/2up

Hyvärinen, M. (2006). Towards a conceptual history of narrative. In M. Hyvärinen, A. Korhonen, & J. Mykkänen (Eds.) *The travelling concept of narrative,* p. 20-41. Retrieved August 11, 2011 from http://www.helsinki.fi/collegium/e-series/volumes/volume_1/index.htm

Johnson, T. H. (2007, September). The Taliban insurgency and an analysis of *shabnamah* (night letters). *Small Wars and Insurgencies, 18*(3), 317-344.

Labov, W., & Waletzky, J. (1966). Narrative analysis: Oral versions of personal narrative. In J. Helm (Ed.), *Essays on the Verbal and Visual Arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society* (pp. 12-44).

Langellier, K. M. (1989). Personal narratives: Perspectives on theory and research. *Text and Performance Quarterly, 9*, 243-276.

Post, J., Sprinzak, E., & Denny, L. M. (2003, Spring). The terrorists in their own words: Interviews with 35 incarcerated Middle Eastern terrorists. *Terrorism and Political Violence, 15*(1), 171-184.

Propp, V. (1968). *Morphology of the folktale*. Austin, TX: University of Texas Press.

Reissman, C. K. (2005). Narrative analysis. In N. Kelly, C. Horroks, K. Milnes, B. Roberts, & D. Robinson (Eds.) *Narrative, memory & everyday life*. University of Huddersfield: Huddersfield, p. 1-7.

Smith, C. P. (2000). Content analysis and narrative analysis. In H. T. Reid, C. M. Judd (Eds.) *Handbook of research methods in social and personality psychology*. New York, NY: Cambridge University Press, p. 313-335.

SOCIAL NETWORK ANALYSIS

Social network analysis is a research paradigm that encompasses a perspective on social processes and a collection of methodological tools for examining relational ties between the entities of a group. Those entities could be people, groups of people, organizations, states, or some combination thereof. Entities in a networked system could also be physical objects, such as production resources, or abstract elements, such as the knowledge required to complete a particular task. Taken together, the ties that connect pairs of entities form a network, and analysis of such networks can provide insights into social structures, the formation and dissolution of relationships between entities, and the various ways that social influence might be transmitted through the network. Through its focus on entities and their relationships, social network analysis can be characterized as "a science of interactions" (Maoz, 2011). Social network analytic methods have been applied extensively in the social sciences, exploring such issues as the impact of peers on adolescent behavior, the effect of cross-linked boards on business performance, and the roles played by alliance and trade networks in international relations.

METHODOLOGICAL OVERVIEW

Social network analysis comprises a diverse collection of tools and methods that have been developed for the purpose of examining the relationships among social entities, the patterns of such relationships within and between groups, the dynamics of the relationships, and the implications of relational networks on individual attitudes, beliefs, and behaviors, as well as group or organizational performance. Although the roots of social network analysis can be traced to the 1930s, when Jacob Levy Moreno developed sociograms and introduced the field of sociometry, advances in social network methods, in terms of both data collection and analysis, have accelerated in recent years as researchers in many fields have begun to adopt a social-network perspective. This perspective highlights the following key ideas:

- Actors and their actions are viewed as interdependent, with one actor affecting and being affected by another, rather than independent and autonomous.
- The relational ties that connect actors are channels for the transfer or exchange of material resources (e.g., money, equipment) and nonmaterial resources (e.g., information, attitudes).
- Network models conceptualize structure (e.g., social, economic, political) in terms of the patterns of relations among entities.
- Network models that focus on individuals view structure as providing both opportunities for and constraints on individual action.

The focus on the *relationships among units* is the fundamental feature of social network analysis that distinguishes it from other social science research methods. In other words, social network analysis applies to collections of entities and the ties that connect those entities. For example, while traditional survey research typically focuses on the responses provided by a sample of disconnected individuals, social network research requires additional information about how those individuals are related. The subsequent analysis of social network survey data would examine not only how individuals responded to particular survey questions about their attitudes and actions, but also how the nature of people's responses were associated with the patterns and characteristics of their relationships. Two possible associations might be investigated. First, an investigator might try to determine whether the attitudes people reveal in their survey responses affect the relationships they form with others. Second, an investigator might try to determine whether respondents' existing relationships affect the attitudes they express in the survey. These alternatives, which can be characterized as *selection* and *influence*, respectively, are just two of many research questions that can be investigated using social network methods.

WHY USE SOCIAL NETWORK ANALYSIS?

Social network analysis methods are applicable to any investigation in which the problem hinges on the relationships among entities. Social network analysis has been applied extensively across the social sciences, and Maoz (2011) makes a strong case for employing social network analysis in the study of international relations, a domain in which he believes the approach has been underutilized. While introducing his general theory of networked international politics (NIP), Maoz writes:

> International relations have evolved as a set of interrelated cooperative and conflictual networks. These networks coevolve in constant interaction with each other, and this interaction has important implications for the behavior of nations and for the structure of the international system. To understand where we were nearly two hundred years ago, how we got from the end of the Napoleonic Wars to the hierarchical system of the present, and where we might go in the future, we must understand how these networks were formed, how they change, how they affect each other, and how they condition the behavior of state and non-state units. (p. KINDLE)

Social network analysis provides a framework for defining problems in terms of actors and the relations that connect them, a variety of techniques for acquiring and organizing relational data, and a set of tools for analyzing such data and modeling interconnected social systems.

FUNDAMENTAL CONCEPTS

The following list of key concepts was adapted from Wasserman & Faust (1994).

> **Actors.** Actor is a term for the entities that make up a social network, though the term should not be taken to imply that network entities have the will or capacity to act. Actors are discrete social units, such as individuals, groups, companies, or nations. Generally, actors coexist within a physical or conceptual boundary. Thus, a network could comprise individuals in a group, departments in a corporation, companies in an industry, or member states of an international trade agreement. One-mode networks include actors that are all

of the same type, while multi-mode networks include actors from different types, levels, or sets.

Relational tie. Actors are linked together by social ties. Such ties can represent any of a wide variety of relationships, including friendship, kinship, economic exchange, mutual affiliation with a group, or a physical connection, such as a road between two cities or a path between two homes.

Dyad. A relational tie connects two actors and is best thought of as a property of the pair, rather than as a feature of either one of the two actors. A dyad is a unit of analysis comprising a pair of actors and the (possible) tie(s) between them. Certain network analyses focus on dyads and issues, such as reciprocity (whether or not a tie from actor *i* to actor *j* is matched with a tie from actor *j* to actor *i*) and multiplexity (whether certain kinds of relationships tend to appear together, such as friendship ties and advice-seeking ties).

Triad. A triad is a unit of analysis comprising three actors and the (possible) tie(s) between them. Extending the unit of analysis from a dyad to a triad introduces new opportunities for study, such as transitivity (whether a tie from actor *i* to actor *j* and a tie from actor *j* to actor *k* predicts a tie from actor *i* to actor *k*) and balance (if actors *i* and *j* 'like' each other, then they should be similar in their evaluation of a third actor, *k*; however, if *i* and *j* 'dislike' each other, then they should differ in their evaluation of *k*).

Subgroup. Subgroups are related to dyads and triads, referring to any subset of actors and the ties between them. An important area of social network research involves defining criteria and analytic methods for identifying subgroups of actors.

Group. In social network analysis, the group is the full collection of actors on which social ties are measured. An investigator must establish the criteria (theoretical, empirical, or conceptual) for establishing the boundaries of a group and for justifying that a particular finite set of actors belong together in that group. That a group consists of a finite set of actors is primarily an analytic restriction, although network analyses can be conducted on extremely large groups, and the focus generally should be on establishing meaningful group boundaries.

Relation. A relation is the collection of a specific kind of ties among the actors in a group. Relations can be characterized as a set of rules that define whether, how, and to what extent any given pair of actors is connected. In a relational network, the rules define a connection between two actors. Many relations can potentially be defined for a given group, such as friendship and professional affiliation among a group of individuals, or diplomacy and trade among a group of nations. In an affiliation network, the rules define a connection between an actor and an event, organization, or group.

Social network. Given the above terms, a social network can be defined as a group of actors and the relations defined on them. The presence of relations is thus a critical and defining feature of social networks.

REQUIREMENTS

DATA

High (4)

Analyses of real networks are quite data intensive, and considerable time and resources are often expended on data collection. The collection of social network data is challenging. Techniques that have been employed for collecting social network data include surveys/questionnaires, interviews, observation, archival records, experiments, and diaries. Although the data requirements for simple investigations, such as an analysis of the egocentric networks of a small, well-defined population of individuals, may be satisfied relatively quickly and at relatively low cost, investigations of network phenomena that span all individuals within a social system, such as tracking the spread of diseases or ideas, may require multiple waves of resource-intensive data collection over an extended period of time.

SET UP TIME

Very Long (5)

The acquisition and management of social network data is a challenging process. Regardless of the data collection method, considerable time is often necessary to plan and execute network data collection. One of the more common methods for acquiring data from a small social network is by using surveys. Research has shown that the design of a network data collection survey can have considerable impact on the quality of network data obtained. Once a survey has been designed, tested, and validated as a useful data collection instrument, the survey must be administered to all individuals in the social system under study. Once that process is complete, time must be scheduled for processing the acquired data and preparing it for analysis. Finally, network data often contain considerable information that could be revealing or damaging if made public, care must be taken and resources must be committed to protect the privacy of participants.

EXECUTION TIME

Short (0.25)

The time required to analyze social network data can vary considerably depending on the particular analysis being performed. For basic computations of descriptive network parameters, such as the distribution of network degree values over all nodes in a network, execution time is typically short even for very large networks. For more sophisticated analyses, such as modeling the probability of observing a tie between two nodes based on individual characteristics of each of those nodes and joint characteristics of pairs of nodes, computation requirements can be very high and execution times can be very long—up to 24 hours depending on computing power—even for relatively small networks.

SKILL SET/EXPERTISE

Specialized-average (1)

A variety of software tools exist for performing basic analyses of social network data and for preparing visualizations of social networks (also known as sociograms). Many basic analyses therefore do not require advanced training in network methods. However, planning and executing a study such that the investigation is grounded in sound social theory and the network data are both reliable and valid for the intended analysis can be challenging and may require specialized social network analysis skills.

As noted previously, the planning and execution of a theoretically grounded study of social networks requires considerable expertise. Moreover, while some basic descriptive statistics can be produced from social network data without extensive training, a research with expertise in social network analysis methods and tools will be in a much better position to determine which measures should be computed for a particular analysis and how the study—including its design, execution, and data analyses—can best be executed so as to produce results that are valid and reliable.

TOOLS

Specialized (2)

Social network analysis is often facilitated by tools that have been developed to handle network data and perform specific analyses. While some such tools are costly, there are many tools that are currently available at virtually no cost to the user. For example, several packages for conducting network analyses ranging from very basic to highly sophisticated have been written for the open source statistical analysis package R. Note that tools that are freely available often tend to require a higher degree of end user expertise.

COST

$$$ (12.25)

STRENGTHS AND WEAKNESSES OF SOCIAL NETWORK ANALYSIS

- The ideal method when investigating questions about the role of social structure
- Good network data are difficult and expensive to acquire
- Missing data, memory recall problems, roster completeness
- Network data go stale, sometimes very quickly

PROBLEMS THAT CAN BE ADDRESSED WITH SNA

Researchers in different disciplines of the social and behavioral sciences have used social network analysis to examine a variety of topics, including:

- How are an individual's personal and professional networks related to occupational mobility?
- How is individual well-being affected by urbanization and its consequent impact on the structure of personal relationships?
- How do political and trade alliances among nations affect the global economic system?
- How do the relations among members of a group affect the efficiency and accuracy of group problem solving?
- How do the relations among individuals affect the diffusion and adoption of technological innovations?
- What factors affect the formation of coalitions among groups of actors?

In the context of international relations, Maoz (2011, italics in original) applies social network analysis to investigate the following questions:

- How, why, and when do different international networks form?
- How do they change over time? What factors determine the nature, magnitude, and types of change in a given network?
- How do different networks affect each other? Do changes in one network affect changes in the structure or characteristics of other networks? If so, how do cross-network relations work, and what are their consequences?
- How do the structure and characteristics of international networks affect various historical processes, such as changing levels of international stability, the degree of economic inequality, and transformations in the structure of the system?
- What are the relationships between *nondiscretionary* networks (e.g., geographic or cultural networks) and *discretionary* ones (e.g., alliances, trade, international organizations)?

Briefly stated, social network analysis methods are applicable when investigating problems that involve social relations among actors, involving questions about both the formation (and dissolution) of relational ties, as well as the ways in which relational ties affect actors' attitudes and behaviors and conversely.

RELATED IDEAS

- Complexity and complex adaptive systems and simulations
- Agent-based systems and simulations
- Game theory (esp. in network-based simulations)

CONNECTION TO CANS

Dynamic networks of interacting states—the actions of one state cannot be completely isolated from the structure of relational ties that state has with other states…

FURTHER READING

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek.* New York, NY, USA: Cambridge University Press.

Maoz, Z. (2011). *Networks of Nations: The Evolution, Structure, an Impact of International Networks, 1816–2001.* New York, NY, USA: Cambridge University Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* New York, NY, USA: Cambridge University Press.

STATISTICAL ANALYSIS & MODELING

Statistical analysis provides a method of examining a question of interest in a systematic manner through the collection, analysis, and interpretation of quantitative data; data that is expressed in numerical form and is replicable. One of the central roles statistical analysis plays is that of description. In descriptive research, we can examine a phenomenon of interest to determine answers to questions such as "How many terrorist attacks occurred in 2010?" Statistical analysis can also help researchers identify explanations for phenomena. Explanatory research seeks to identify causes or effects of phenomena. For example, researchers interested in terrorism might ask, "What factors are related to an increase in terrorist attacks?" Finally, statistical analysis can help evaluate the implementation and effectiveness of programs and policies. Evaluation research can provide answers to questions such as "Have the increased security measures put in place in US airports post 9/11 been effective in reducing the threat of terrorist attacks?"

METHODOLOGICAL OVERVIEW

WHAT IS STATISTICAL ANALYSIS?

Observations about persons, things, and events are central to answering questions about human behavior, state behavior, and even institutional behavior. Statistical analysis is the scientific process of converting these observations to quantitative (numerical) data and examining, summarizing, and interpreting these observations to assess relationships, patterns, trends, or causes. Certain information is, by nature, numerical. For example, it is easy to see how a person's age, height, or the temperature is quantitative data. But other observations may also be given numerical values. For example, observations about political party affiliation, gender, religious affiliation, etc. may be given numerical values, allowing qualitative (non-numerical) data to be used in statistical analysis.

The general purpose of statistical analysis is to provide meaning to what otherwise would be a collection of numbers. The meaningfulness of the data is a product of the clarity with which one specifies the problem or questions being addressed and the precision with which the pertinent information is gathered and analyzed.

The steps involved in a statistical analysis are:

1. Deciding on the questions you wish to explore
2. Proposing some reasonable explanations which have the potential to answer these questions
3. Developing conceptual and operational definitions for your concepts
4. Collecting appropriate data

5. Performing the appropriate statistical analysis to answer your question

Properly designed statistical models minimize subjectivity and allow a researcher to summarize their findings in a clear, precise and reliable manner, as well as providing a means to assess the statistical confidence – the accuracy – of their conclusions. Statistical analysis requires an understanding of statistical assumptions and the limitations of the various models. Phenomena that are difficult to conceptualize or quantify can be problematic in a statistical analysis. While statistical analysis can provide generalizable findings, this type of analysis loses the detail that is retained in other types of analyses (i.e., case studies).

WHY USE STATISTICAL ANALYSIS?

There are numerous reasons to use statistical analysis. First, it can be used to process large amounts of data and summarize that data in a shorter form. For example, you might have conducted a survey of 500 people in an attempt to understand viewpoints on defense spending. In the survey, you might have asked questions about their level of education, their gender, their political affiliation, and whether they approved of the amount of money being spent on defense. This volume of data needs to be organized in some logical manner; otherwise, it is unusable. Statistical analysis can be used to give you a snapshot of the data. For example, you could determine how many males answered the survey and how many females answered the survey. You could determine whether males were more likely to be Republicans than females, and you could determine whether females were less likely to approve of the amount of money being spent on defense than males—information that would not be discernable just by looking at the surveys. Additionally, statistical analysis can be used to understand some process and possibly predict based on the data collected. Statistical analysis can also provide a way to objectively report on how unusual an event is based on historical data. Finally, statistical analysis gives us a way to quantify the confidence we can have in our inferences.

HOW TO CONDUCT A STATISTICAL ANALYSIS

Conducting a successful statistical analysis requires thought and preparation. Carefully designed research questions, thoughtful consideration of conceptual and operational definitions, collection of the appropriate data, and knowledge of the correct statistical method to employ is imperative.

Problem Definition

Carefully defining one's problem via specific research questions or hypotheses provides a foundation for a successful statistical analysis. Carefully worded **research questions** or **hypotheses** guide the research by defining which concepts must be defined, which variables will be scrutinized, and what relationships may exist between or among these variables. For example, the question "Are women more liberal than men" might, at first glance, seem like a clear research question; but, as you will see in the next section, the concept "liberal" needs more clarification.

Conceptualization

Conceptualization is the process by which we specify what we mean when we use particular terms. The same concept can, and often does, refer to a variety of different concrete terms. "Are women more liberal than men?" The answer depends on what you mean by *liberal.* Do you mean *liberal* in terms of environmental protection, or *liberal* in terms of gun control? *Liberal* may mean support for gun control for some. For others, the concept might refer to support for environmental protection.

A **conceptual definition** clearly describes the concept's measurable properties and specifies the **unit of analysis** (the entity - people, nations, states, etc. we want to describe and analyze) to which the concept applies. Using the example above, a researcher might frame a conceptual definition of liberal as follows: Liberal is the extent to which individuals support gun control. This statement clarifies a vague idea, liberal, by making reference to a measurable attribute – support for gun control. Note the phrase "the extent to which." This phrase suggests that the concept's measurable attribute – support for gun control – varies across individuals. Thus, someone who supports gun control is "more liberal" than someone who does not support gun control. The conceptual definition also makes it clear that this definition applies to individuals.

Properly constructed conceptual definitions communicate three things:

1. The variation within a measurable characteristic or set of characteristics;
2. The subjects or groups to which the concept applies;
3. How the characteristic is to be measured

Following is a workable template for stating a conceptual definition that meets all three requirements:

The concept of ____________________ is defined as the extent to which __________ exhibit the characteristic of ____________________" (Pollock 2009:12)

We could use this template to develop a conceptual definition for the term *liberal* as follows:

The concept of <u>liberal</u> is defined as the extent to which <u>individuals</u> exhibit the characteristic of <u>supporting gun control.</u>
Clear conceptual definitions make it easier to operationalize and accurately measure concepts. In our example, we used support for gun control as a conceptual definition for liberal. Thus, our research question moves from the vague question of "Are women more liberal than men?" to the question of "Are women more likely to support gun control than men?" This clear, precise research question can be answered by statistical analysis as we are able to operationalize the concept "support for gun control."

Operationalization

Properly constructed conceptual definitions enable the researcher to develop specific **measurements** for their concepts, minimizing validity and reliability issues. When a researcher specifies the operational definition of a concept, the precise meaning of that concept becomes clear. Now it is necessary to go one step further and specify how we are going to measure that concept. It is useful to think of **operationalization** as the final stage in the process of defining your concept.

Using the example from the previous section, the task now is to determine how we can measure whether someone supports gun control – to **operationalize** our concept. The researcher might decide to operationalize support for gun control by using a public opinion poll question – asking individuals if they agree with the following statement: "All individuals who wish to purchase any type of firearm should be required to pass a background check" – a question that asks for a simple yes or no response. Alternatively, the researcher might use the same question, but ask respondents to indicate if they "strongly agree, agree, do not have an opinion, disagree, or strongly disagree" with the statement. Finally, the researcher might ask respondents to choose a point on a 100 point scale (0-100) that indicates their support for gun control (0=does not support any gun control; 100=supports maximum amount of gun control). Any of these would be acceptable operational definitions.

The abstract concept of liberal has now been given an operational definition that can be used to measure that concept for individuals. The operational definition indicates precisely what observations need to be made. As noted above, there are many possible operational definitions for a concept. The important thing is to think carefully about the operational definition you choose and ensure that the definition coincides closely with the meaning of the original concept. The establishment of clear and unambiguous operational definitions is a crucial step for both data collection and analysis.

Variables and their Measurement

The operationalization of a concept provides a blueprint for its measurement. **Attributes** are characteristics or qualities that describe some object such as a person, institution, country, etc. **Variables** – metrics - are logical groups of these attributes.

All variables share certain features. Every variable has one name and at least two values or attributes – there is variation. The attributes may vary by category, such as gender (male, female) or religion (Catholic, Protestant, Jewish, etc.), or the attributes may vary in magnitude, duration and/or intensity, such as "age" where the attributes vary in magnitude, "months" where the attributes vary in duration, or "approval" where the attributes vary in intensity.

Variables may be categorized in a number of ways, the first being the role they play in the research question. **Dependent variables** represent the concept we are trying to explain. In our example, we are trying to explain support for gun control; this would be our dependent variable. **Independent variables** are those that we believe explains the variation in the dependent variable. In our example, gender is what we think explains the variation in support for gun control. We are hypothesizing that women are more likely to support gun control than men are.

We might have some additional variables that we also believe explain the variation in support for gun control. **Intervening variables** represent factors that might alter the relationship between the independent and dependent variables. For example, even if gender does explain the variation in the dependent variable, are there other possible factors that might interact with gender? Do women with higher incomes support gun control more than women with lower incomes? Or, is it possible that men with higher incomes support gun control more than women with lower incomes, but not more than women with higher incomes? Does education alter the relationship between gender and support for gun control? Intervening variables must also be considered when developing a statistical model in order to ensure the accuracy of your results.

Once categorized according to the above scheme (independent, dependent, intervening), variables must be measured. This is the basis for data gathering. Data gathering employs measurement scales or sets of rules for quantifying and assigning values to a particular variable. The level of measurement involves the type of information our measurements contain and the type of comparisons that can be made across a number of observations on the same variable.

There are three different **levels of measurement**: nominal, ordinal, and interval/ratio[23]. Very few concepts inherently require a particular level of measurement, so the level used in any specific research project is a function of the imagination and resources of the researcher and the decisions made when the method of measuring each of the variables is developed.

The term nominal means to name. Therefore, a **nominal scale** does not measure but rather names. A nominal variable thus consists of named categories. Gender is an example of a nominal variable. The attributes of a nominal variable have no inherent order. For example, gender is a nominal variable in that being male is neither better nor worse than being female. Persons, things, and events characterized by a nominal variable are not ranked or ordered by the attributes. For purposes of data analysis, we can assign numbers to the attributes of a nominal variable but must remember that the numbers are just labels and must not be interpreted as conveying the order or value of the attributes.

In our earlier discussion of operationalization, we suggested that one way of operationalizing our concept was to ask individuals if they agreed with the following statement: "All individuals who wish to purchase any type of firearm should be required to pass a background check"—a question that asks for a simple yes or no response. This would result in a nominal dependent variable, a variable whose attributes have no inherent order.

[23] There are some important differences between interval and ratio level data, particularly when dealing with some advanced statistical methods, but in most statistical analyses, interval and ratio data are considered equivalent.

The term ordinal means to order. In other words, an **ordinal scale** is a rank ordering of attributes, with a categorization in terms of more than or less than. Examples of variables measured on an ordinal scale would be pain levels (on a high, medium, or low scale), or the rank ordering of attitudes toward a specific policy (on a highly disapprove, moderately disapprove, indifferent, moderately approve, or approve scale). Although the ordinal level of measurement yields a ranking of attributes, no assumptions are made about the "distance" between the classifications. We do not assume that the difference between persons who greatly approve of a program offering and ones who moderately approve is the same as the difference between persons who moderately approve of the program and ones who are indifferent to it. For data analysis purposes, numbers are assigned to the attributes (i.e., 1=highly disapprove, 2=moderately disapprove, 3=indifferent, 4=moderately approve, 5=highly approve, but the numbers are understood to indicate rank order only and the "distance" between the numbers has no meaning.

In our earlier discussion of operationalization, we suggested that one way of operationalizing our concept was to ask individuals if they agreed with the following statement: "All individuals who wish to purchase any type of firearm should be required to pass a background check," but rather than indicating yes or no, respondents would be asked to indicate if they "strongly agree, agree, do not have an opinion, disagree, or strongly disagree" with the statement. This operational definition would result in an ordinal dependent variable, a variable whose attributes have a clear rank order.

Interval/Ratio measurements not only tell the order of things, they also measure the distance between values. For instance, assume you measure two patients' temperatures, one as 98 degrees and one as 100 degrees. Not only does the second patient have a higher temperature than the first, but their temperature is 2 degrees higher. The attributes of an interval/ratio variable are assumed to be equally spaced. For example, temperature on the Fahrenheit scale is an interval variable. The difference between a temperature of 45 degrees and 46 degrees is taken to be the same as the difference between 90 degrees and 91 degrees.

In our earlier discussion of operationalization, we suggested that one way of operationalizing our concept was to ask respondents to choose a point on a 100-point scale (0-100) that indicates their view on the importance of gun control. This operational definition would result in an interval/ratio level dependent variable, a variable where the distance between the numbers is considered equal. For example, if one respondent had chosen 46, and another had chosen 48, that 2-point difference would be considered equivalent to the same 2-point difference if one respondent had chosen 87 and another had chosen 89.

Whenever possible, data should be gathered at the highest level possible – the interval/ratio level (or the highest level appropriate for your concept; if you want a yes/no answer, you should collect the nominal level data). The higher level of precision provided by interval/ratio level data allows for more powerful statistical testing. Moreover, high-level data easily can be converted to lower levels, i.e. ordinal or nominal. The reverse is not true. Using our example of asking a respondent to choose a point on a 100-point scale that indicated their view on the importance of gun control (obtaining interval/ratio level data), that data could be converted to ordinal data by dividing the

100 point scale into 5 (or however many ordered categories you wish) equivalent categories. For example, scores ranging from 0-20 could be considered "strongly disagree," scores ranging from 21-40 could be considered "disagree," scores ranging from 41-60 could be considered "do not have an opinion," scores ranging from 61-80 could be considered "agree," and scores from 81-100 considered "strongly agree." This would convert the interval/ratio level data to ordinal data. You could also choose to convert the interval/ratio level data to nominal data by simply considering all scores from 0-50 and being anti-gun control and scores from 51-100 as being pro gun control.

Lower-level data (nominal and ordinal) can never be converted to higher-level data. If there is any doubt as to whether you might need higher-level data, it is suggested that data be gathered at the highest possible level.
What statistical analysis do you use?

The type of statistical analysis you use will be determined by the type of question you wish to answer and the type of data you have or can gather. Depending on the type of question you want to answer, you may rely either on descriptive statistics or inferential statistics for your answer, or both.

Descriptive statistics summarize vast amounts of data and information in an organized manner. Descriptive statistics may be used when it is not desirable to develop complex research models and can provide preliminary information before undertaking more advanced statistical analysis. **Associative statistics** seek to identify meaningful relationships, and inferential statistics are used to understand process and possibly predict future behavior or events. Additionally, associative and **inferential statistics** provide a way to quantify the confidence we have in our inferences.

Table 1. Types of Statistical Analysis

Type of Analysis	Model	Type of question that can be answered	Dependent Variable	Independent Variables	Observations
Descriptive	Mean Median	What is the average income for Community A?	Interval Ratio	Interval Ratio	Snap shot Time Series
Descriptive	Mode	How many people are in favor of gun control?	Nominal	Nominal	Snap shot Time Series
Descriptive	Mode	Are we spending too much, too little, or about the right amount on education?	Ordinal	Ordinal	Snap shot Time Series
Associative/ Inferential	Crosstab	Are women less likely than men to support the death penalty?	Nominal Ordinal	Nominal Ordinal	Snap shot

Type of Analysis	Model	Type of question that can be answered	Dependent Variable	Independent Variables	Observations
Associative/ Inferential	Correlation	Do states with lower levels of education have fewer women legislators than states with higher levels of education residents?	Interval Ratio	Interval Ratio	Snap shot
Inferential	One-Way ANOVA	Do women contribute more money to political campaigns than men?	Nominal Ordinal	Interval Ratio	Snap shot
Inferential	Two-Way ANOVA	Do states with mandatory automobile inspections and lower population densities have fewer auto fatalities than those without mandatory automobile inspections and higher population densities?	Interval Ratio	Nominal Ordinal	Snap shot
Inferential	Single Sample *t* test	Is New York State's average income lower than the national average?	Nominal	Interval Ratio	Snap shot
Inferential	Paired Sample *t* test	Did confidence levels in government increase after the election?	Nominal	Interval Ratio	Time Series
Inferential	Independent Sample *t* test	Will men give the Republican Party higher acceptance ratings than women?	Nominal	Interval Ratio	Snap shot

Type of Analysis	Model	Type of question that can be answered	Dependent Variable	Independent Variables	Observations
Inferential	Multiple Regression	Is there a relationship between a person's income, their political affiliation, and the number of times a person contacts their Senator?	Interval Ratio	Nominal Ordinal Interval Ratio	Snap shot Time Series
Inferential	Logistic Regression	Why do people in the United State vote or not vote?	Nominal Ordinal	Nominal Ordinal Interval Ratio	Snap shot Time Series
Inferential	Time Series	Has Country A become more democratic since 1970?	Interval Ratio	Nominal Ordinal Interval Ratio	Time Series
Inferential	Hazard	When will a war occur?	Nominal Ordinal	Nominal Ordinal Interval Ratio	Time Series

Statistical Significance

Whereas descriptive statistics simply portray data, associative and inferential statistics test the likelihood or probability of a set of observations relative to chance. For this reason, associative and inferential statistical procedures provide a level of probability or "*p"* value. The ***p* value** represents the probability that the observed findings are a "chance" occurrence, i.e., due to random fluctuations or errors in sampling. A *p* value of .01, therefore, indicates that the probability is 1 out of 100 that the observed finding is a chance event. Conversely, one could say with 99% confidence that the observed finding is significant, there is a relationship between the variables, the findings did not occur simply by chance.

The acceptable *p* value in the social sciences is .05. Should the results of a statistical analysis provide a value greater than .05 (e.g., $p = .10$), the researcher would not be willing to claim that the findings were meaningful. In other words, the findings would not be statistically significant. The probability of the findings being a chance occurrence would be too high to have confidence in the results.

Using our example, we developed a model with "support for gun control" as our dependent variable and our primary independent variable as "gender." We believe there might be some intervening variables, for example, income and education. Let us say that when we perform our statistical analysis, we find that the *p* value associated with gender equals .04, the *p* value associated with income equals .08, and the *p* value associated with education equals .05. These *p* values determine which of these variables are statistically significant predictors of our dependent variable, support for gun control. In this case, gender is a statistically significant predictor as the *p* value is less than or equal to .05, as is education. Income is not a statistically significant predictor, as the *p* value of .08 is higher than .05.

REQUIREMENTS

DATA

High (4)

As mentioned earlier, one of the advantages of statistical analysis is its usefulness in summarizing large amounts of data. That is also one of the disadvantages of statistical analysis, the need for large amounts of data. While some statistical analyses can be conducted with a small amount of data, inferential analyses require the collection of a significant amount of data. There is no magical amount of data that you must have to conduct a statistical analysis. The more complex your model (the more independent variables), the more data you will need.

SET UP TIME

Average (3)

The most time-consuming part of the analysis is usually the data collection. Even if information is readily available to the researcher, it takes time to convert that information to useable (numerical) data that can then be used in a statistical analysis. In many cases, it is necessary to gather the data necessary; an exercise that may include conducting public opinion polls, analyzing existing sources of data (i.e., the U.S. Census data), or conducting interviews.

EXECUTION TIME

Short (0.25)

Once the data required for a research question is collected, the time needed to run analyses is short.

SKILL SET/EXPERTISE

Specialized-average (1)

Statistical analysis requires training in statistical methodology. While many people have been trained in the use of statistical analysis programs, it is not simply a matter of plugging in numbers and running a program. Accurate and useful statistical analysis is conducted by individuals trained in the development of research questions and hypotheses, as well as the collection of appropriate data and statistical theory. Additional training in social science research design and in statistical methodology is required.

TOOLS

Average (1)

Statistical software programs such as SPSS, SAS, STATA are required for the analysis data.

COST
$$ (9.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

See Table 1. Types of Statistical Analysis for an overview of what types of questions could be answered with statistical analysis.

FURTHER RESOURCES

Giles, D. 2002. *Advanced Research Methods in Psychology*. New York: Routledge.

Gravetter, F. and L. Forzano. 2009. *Research Methods for the Behavioral Sciences.* New York: Wadsworth, Cengage Learning.

Pollock, P. 2009. *The Essentials of Political Analysis: Third Edition*. Washington, D.C.: CQ Press.

Spicer, J. 2004. *Making Sense of Multivariate Data Analysis*. New York: Sage Publications, Inc.

StatSoft Electronic Statistics Textbook. http://www.statsoft.com/textbook/basic-statistics/

Tabachnick, B.G and I.S. Fidell. (1996) *Using Multivariate Statistics: Third Edition*. New York: Harper Collins.

Weinberg, S. and S. Abramowitz. 2008. *Statistics Using SPSS: An Integrative Approach*. New York: Cambridge University Press.

STRUCTURED ARGUMENTATION

Structured Argumentation refers to techniques for organizing, communicating or evaluating lines of reasoning. In this report, we describe first what we call Analytic Structured Argumentation, which couples formal, explicit elucidation of lines of reasoning with the decomposition of complex propositions into levels of constituent subordinate propositions with the goal of reaching elemental claims that can be most readily evaluated in light of available supporting information or expertise. A second type of structured argumentation, developed from the work of Stephen Toulmin begun in the 1950s, focuses on making explicit the claim made by an argument and the evidence offered to support it so that the nature and validity of the argument may be made plain. This latter type of structured argumentation is not the focus of this digest; however, it will be covered briefly, and its connection to the analytic process described will be discussed.

METHODOLOGICAL OVERVIEW

ANALYTIC STRUCTURED ARGUMENTATION

The goal of Analytic Structured Argumentation is to decompose higher-order, complex propositions into constituent propositions, with the analysis ending at basic claims that can be directly assessed by available information in the form of data or expertise. Propositions are linked to the subordinate propositions that directly imply it by rules of inference that dictate the conditions under and the degree to which the sub-propositions combine to imply the truth of the superordinate proposition. The evaluation of the veracity of the top-level proposition percolates through this hierarchy from the elemental claims through intervening propositions in accordance with the rules of inference specified in the structure of the argument.

It is convenient and informative to identify such a structured argument with the graphical object called a tree. An example of such is a diagram given in Figure 4.

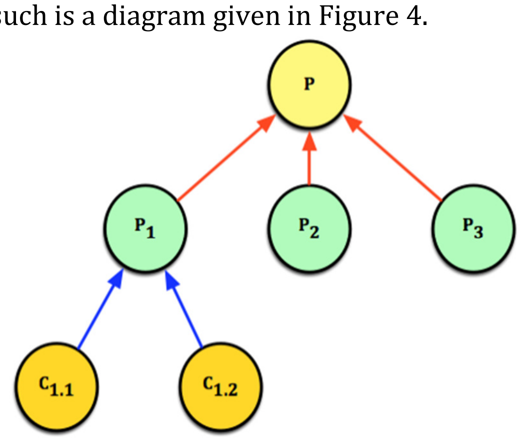


Figure 4. Example argument tree

In keeping with the nomenclature of tree diagrams such as this one, the top-level proposition is often referred to as the root, and the bottom level claims are called leaves. Note that in Figure 4 we have suppressed displaying the claims that underlie P_2 and P_3 to make the diagram more readable for the immediate purpose. Note, too, that we have colored the edges on the two levels differently to indicate that the inference rules need not be uniform across the structure.

Structuring arguments analytically in this fashion comprises three interrelated requirements:

- The analysis of higher-order propositions into simpler constituent propositions, eventually down to what we have described here as elemental claims.
- The specification of inference rules for each "family" unit. There is no requirement that the rules be uniform across the argument structure, and it may be entirely necessary that they not be.
- Assessments about the degree of veracity of each of the elemental claims in the structure. This is often expressed on a discrete ordinal scale, with the bottom of the scale an assessment that the claim is very likely untrue and the top an assessment that it is very likely true.

As a concrete example, suppose that

P is the proposition: "Bob will receive a raise next year,"
P_1 is the proposition: "Bob will receive an outstanding performance evaluation,"
P_2 is the proposition: "Bob will be recruited by competitors," and
P_3 is the proposition: "Bob cannot be readily replaced."

As before, we imagine, but will not detail, claims undergirding P_2 and P_3 and focus on the claims undergirding P_1. Suppose that:

$C_{1.1}$ is the claim: "Bob has received extremely positive feedback from his superiors throughout the year," and
$C_{1.2}$ is the claim: "Clients have been extremely pleased with the work Bob has produced on their behalf this year."

Plausible inference rules for this structure might derive for P_1 a truth value that averages the values of $C_{1.1}$ and $C_{1.2}$ and derives for P the greater of the truth value derived for P_1 (from the claims $C_{1.1}$ and $C_{1.2}$) or the average of the values derived for P_2 and P_3 from their underlying claims (or subordinate propositions). Of course, specifying the inference rules, just as with the decomposition of propositions into their subordinates, is subject to the understanding and judgment of those creating the analysis.

TOULMIN MODEL FOR STRUCTURED ARGUMENTATION

Structured argumentation following the work of Stephen Toulmin focuses on a different aspect from the analytic approach outlined above. It is not concerned with decomposing complex arguments into propositions and their subordinates, but rather with explicit framing of an assertion with the collection of logical elements that both augment and question its credibility. The model is illustrated in Figure 5. Elements shaded in yellow are considered essential to any argument, while those in green may or may not apply to every argument.

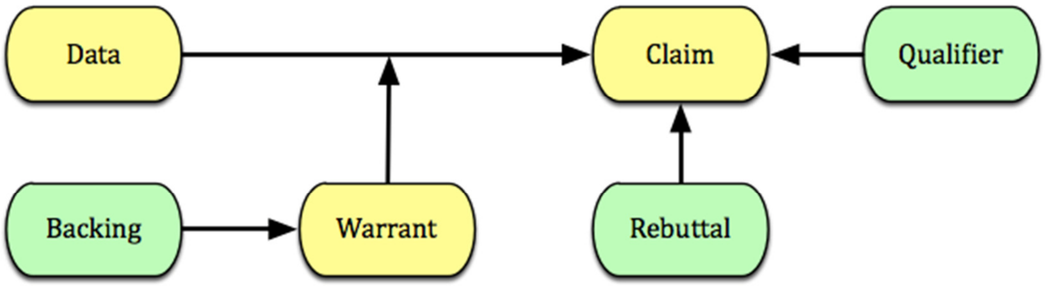


Figure 5. Elements of the Toulmin model for structured argumentation

Elements of a the Toulmin Model

Claim: The central assertion being proposed by the argument that is subject to scrutiny and validation.

Data (or evidence): The facts used to substantiate the claim.

Warrant: The reasoning that explains why the data supports the claim.

Backing: Information that supports the reasoning behind the warrant, which may be required if the warrant itself is not sufficiently convincing to establish the connection between the data and the claim.

Rebuttal: Statements of conditions or circumstances that may restrict the validity of the claim.

Qualifier: Qualifications as to the strength or certainty of the claim. A qualifier would be associated with words or phrases like "probably," "possibly," "almost certainly," etc.

CONNECTION BETWEEN THE TOULMIN MODEL AND ANALYTIC STRUCTURED ARGUMENTATION

The Toulmin model connects to Analytic Structured Argumentation at the level of the basic claims upon which the rest of the analytic structure rests. A Toulmin-type structuring of the basis for these claims and the evidence supporting them can assist analysts in assigning truth values and communicating the reasoning behind these assignments.

REQUIREMENTS

Analytic Structured Argumentation, by its very nature, demands problems that can be framed as a hierarchical proposition structure terminating in claims amenable to direct evaluation.

DATA

Average (3)

Claims at the base of the analytic structured argumente must be substantiated through expertise or data. The extent to which the latter is required will depend upon the number of elemental claims appearing in the analysis, the nature of the data available to evaluate these, and the number of distinct scenarios requiring independent assessment of the elemental claims.

SET UP TIME

Average (3)

The time required creating the structure and assessing the elemental claims may be considerable.

EXECUTION TIME

Average (0.5)

Time expended creating the structure can be considered a one-time start up cost. To the extent that the use of the analysis is to study alternative scenarios across which the elemental claims may have different levels of likelihood, it may be necessary to allocate considerable time to the assessment of these claims for each scenario.

SKILL SET/EXPERTISE

Specialized-average (1)

The skills required to use an analytically structured argument are not highly specialized once the structure is created. The greatest degree of difficulty for the end user arises in the assessment of the elemental claims. This should be within the skill set of the general analyst.

Creation of the analytic structure, that is, decomposing propositions into their subordinates and crafting rules of inference to move upward through the structure, can be a highly rarefied skill. As this is created, once used repeatedly, situations for each argument decomposition, outside expertise can be brought to bear on this aspect of the analysis.

TOOLS

Minimal (1)

The primary resources required to construct an analytically structured argument are the expertise to create the structure and the data or expertise to assess the elemental claims. Argument trees are a very effective and efficient means of presenting the basic structure of an analytic argument.

COST

$ (8.5)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Structured argumentation can answer the following types of analytic questions:

1. What does an actor perceive as true about the world?
2. What are the key hypotheses in an actor's perceptions of the world?
3. How can influence be brought to bear on those perceptions through providing additional evidence (e.g., of U.S. intentions)?
4. What effect will changes in evidence within a decision environment have on ultimate perceptions?
5. What is the specific causal relationship between pieces of evidence and perceptions?
6. Where should further data collection be focused?

FURTHER RESOURCES

Lowrance, J. D. (2007). Graphical manipulation of evidence in structured arguments. *Oxford Journal of Law, Probability & Risk , 6* (1-4), 225-240.

Lowrance, J. D., Harrison, I. W., & Rodriguez, A. C. (2000). Structured Argumentation for Analysis. *Proceedings of the 12th International Conference on Systems Research, Informatics, and Cybernetics: Focus Symposia on Advances in Computer-Based and Web-Based Collaborative Systems* (pp. 47-57). Baden-Baden: International Conference on Systems Research, Informatics, and Cybernetics.

Toulmin, S. (1958). *Uses of argument.* Cambridge, England: Cambridge University Press.

STRUCTURED CASE STUDIES

Identify the levels, observations, temporal domains, and forms for which this method is most suitable.

Levels

- ☑ **Global**
- ☑ **Regional/multi-actor grouping**
- ☑ **Nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ❑ **Large-N**
- ☑ **Small-N**

Temporal domains

- ☑ **Time-series**: Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Forms

- ❑ **Quantitative**
- ☑ **Qualitative**
- ❑ **Quantitative and qualitative**

INTRODUCTION

The method of structured, focused comparisons is commonly employed in the Social Sciences, both to develop hypotheses and to test causal arguments. The method is principally used in small-N studies that compare a limited number of case studies. The research design is “structured” insofar as each case is examined with standardized research questions and data collection, thereby permitting systematic comparison across cases. The method is “focused” because it deals with only those aspects of the cases that are most relevant to the overriding research question.

Structured cases studies can serve a number of important functions: generating hypotheses, making contextualized comparisons, identifying causal mechanisms in a small universe of cases, and addressing causal complexity. They are also useful for examining outliers or “deviant” cases, as well as building theory when few real-world observations exist. The method has potential pitfalls as well, including case selection bias and difficulty adjudicating between competing explanations. It may or may not be possible to reasonably generalize beyond cases examined in the study. The comparative case study methodology can sometimes have difficulty determining *how much* a variable mattered, as opposed to *whether* and *how* it mattered.

METHODOLOGICAL OVERVIEW

Structured case studies are used throughout the Social Sciences to generate hypotheses and test causal relationships. The key feature of the method is parallel data collection across a limited selection of cases. The ultimate goal is to identify causal mechanisms and build theories that go beyond single events.

Although cases studies can include quantitative variables, the method most commonly incorporates historical narratives that draw upon qualitative variables.

The method is not bound by a particular level of analysis. Comparative case studies in the field of International Relations have been done on a broad range of units of analysis, including international systems, international organizations, nation states, terrorist organizations, and individuals.

Likewise, the method can be used to either compare cases that occur in the same time period or examine one or more cases across time.

HOW IS THIS DONE?

George and Bennett (2005) describe three phases for performing structured, focused case studies.[24] These are each discussed below.

Phase One: The Research Design

In the first phase, the researcher determines the goals and structure of the research project. The phase has five key tasks.

- Specification of the problem and research objective: The research question should be scoped, focused, and well defined. If the topic has already received treatment in the discipline, the researchers should place their argument in the context of—either building upon or in opposition to—existing academic theory and literature.
- Specification of the variables: The researcher should clearly articulate which outcomes (dependent variables), causes (independent variables), and intervening or intermediate variables will be examined.
- Case selection: Although random selection is essential to large-N methods, it is rarely employed when examining a small number of cases. Cases should be selected specifically to show contrast; in order to make causal inferences, the cases must generally include variation in the key variables.

[24] The authors provide examples throughout their overview in Chapters 3-6.

- Describing the variance in variables: The researcher will have to strike a balance between richness and parsimony in determining how to categorize and measure the variables. There is no simple formula for achieving this; too much richness and complexity will inhibit generalizable theory building, whereas too much oversimplification will obscure important considerations and distinctions.

- Formulation of data requirements and general questions: The researcher should specify which data are to be obtained from the cases. Where feasible, it is usually desirable to standardize data requirements such that the cases can be systematically compared and analyzed once the case studies are completed. In reality, researchers will often have to contend with different types of data and data sources across cases. While this can complicate standardized data collection, drawing from a variegated set of sources can also mitigate against the biases embedded in any one source.

Phase Two: Carrying Out the Case Studies

Once Phase One is complete (the research question is defined, the variables are identified, and the research design is established), the researcher will move on to performing the case studies.

- The first step in this process is often to familiarize oneself with the cases. This commonly includes reviewing the existing literature, examining interview data, and speaking with experts.

- The researcher will then engage in the process of original research, adhering to the structured questions and data collection outlined in Phase One. Possible primary and secondary research sources include archival materials, memoirs, autobiographies, oral histories, newspapers, official documents, survey data, and interviews. Trachtenberg (2006) offers a comprehensive guide to using original sources and interpreting historical works.

- Once the data is collected, the difficult task begins of deriving sound explanations for the outcomes in each individual case. These explanations should be consistent with the data and take into account competing explanations.

Phase Three: Drawing the Implications of Case Findings for Theory

If the structured, focused methodology has been performed properly, the cases can then be evaluated comparatively.

- Case studies can be used both to develop theory inductively and to test existing theory. During the former, case research can uncover new variables and causal pathways.

- If the goal is instead to test theory, cases can be used to either weaken or strengthen existing historical explanations. While it is difficult for case studies to entirely falsify a theory (unless the theory is expressed in unconditional terms), they can be used to assess how causal explanations can be expanded, narrowed, or contextualized differently.

- Researchers will often seek to generalize their findings beyond the particular cases examined in the study. There are longstanding debates among methodologists about the degree to which case studies can contribute to generalizable theories. King, Keohane, Verba (1994) and Mahoney (2000) offer strategies for drawing causal inferences from small-N qualitative research.

Common Research Designs for Comparative Case Studies

Four common comparative case study research designs are described below. For each design, an example from the International Relations literature is offered in the footnotes:

- "Most different systems" Design: This design is used when vastly different cases share a similar outcome.[25] If a key independent variable is the same across otherwise disparate cases, it can be identified as a potentially important cause of the shared outcome. This research design is similar to John Stuart Mill's "method of agreement."

- "Most similar systems" Design: In this research design, cases are comparable across all relevant variables, except for one crucial causal variable.[26] In this case, the variation across outcomes is attributed to that crucial variable.

- Analyzing "Deviant cases": Deviant cases are those that include outcomes that do not abide by usual causal explanations.[27] "Deviant cases" are critical for uncovering new variables and developing novel hypotheses. Large-N studies that search for statistical significance tend to obscure the potential importance of deviant cases, which may be of particular interest. This reserves a key role for case study research.

---

[25] For example, see Ember, Carol, Melvin Ember, and Bruce Russett. 1992. "Peace between Participatory Politics." *World Politics* 44: 573-599. In support of Democratic Peace Theory, the authors compare the crisis behavior of modern societies with those of preindustrial societies. Despite their many differences, the authors argue that the common trait of participatory processes, when present, decreased the likelihood of conflict between dyads.

[26] For example, see Ray, James Lee. 1995. *Democracy and International Conflict: An evaluation of the democratic peace proposition*. Columbia: University of South Carolina Press. In support of Democratic Peace Theory, Ray compares the crisis of the Fashoda Crisis between Britain and France with the Spanish-American War. He argues that the autocratic nature of the Spanish regime was the decisive factor in leading to war, whereas the Fashoda Crisis between democracies was resolved peacefully.

[27] For example, see Elman, Miriam Fendius. 1997. *Paths to Peace: Is democracy the answer?* Cambridge, MA: MIT Press. Elman seeks to modify Democratic Peace Theory by examining the deviant case of Finland's war with Great Britain during WWII.

- Examining "Hard cases": Hard or crucial cases help to adjudicate between competing explanations. A hard case is one in which usual or alternative causal explanations strongly predict a particular outcome.[28] This research design is particularly effective if that predicted outcome does not occur, and instead, a new theory with a different causal pathway better explains the case.

REQUIREMENTS

DATA

Low (2)

Data for structured, focused comparisons can be either quantitative or qualitative. Typically, however, the data is qualitative and presented in narrative form. Possible primary and secondary research sources include archival materials, memoirs, autobiographies, oral histories, newspapers, official documents, survey data, and interviews.

SET UP TIME

Average (3)

Data collection for case study research can be time intensive. It frequently requires travel, interviews, and primary research.

EXECUTION TIME

Long (1)

Designing a research agenda, identifying gaps in existing research, and articulating subtle or new causal explanations most often require intimate knowledge of the cases. This tends to increase the amount of necessary research time, because it puts a burden on the researcher to develop expertise on the country or event in question.

SKILL SET/EXPERTISE

Specialized-average (1)

Comparative case studies require a demanding set of methodological skills related to research design, variable conceptualization, data collection, theory building, and theory testing.

Resources may be necessary for interviews with subject matter experts (SMEs), access to primary and secondary materials, and to hire additional researchers to assist with data collection in the field. Funds may also be needed for travel, translators, or other support for field research.

---

[28] For example, see Evangelista, Matthew. 1999. *Unarmed Forces: The transnational movement to end the cold war.* Ithaca, NY: Cornell University Press. In contrast to realist theories, which are commonly used to explain Cold War outcomes, Evangelista examines the role of transnational actors in shaping U.S.-Soviet defense and arms control policies.

TOOLS
Minimal (1)

COST
$ (8)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Structured, focused case studies are used to build hypotheses and test causal arguments. As previously mentioned, comparative case studies are less effective at determining *how much* a variable matters, as opposed to *whether* and *how* it matters.

This has an effect on how case study research should be employed as part of the CANS research agenda. Case studies are not well suited to directly answer the top-level questions identified by the 5D framework. Namely:

- How effective is current U.S. force posture for achieving policy objective?
- What would be the optimal force posture to achieve a specific policy objective?
- What strategy is optimal to achieve the objective?

However, case studies are an extremely effective way to test the causal arguments that substantiate underlying claims about the effectiveness of various force postures and nuclear weapons strategies.

For each constituent part of the 5D framework, case studies can help to develop and test causal mechanisms. For example:

Policy Objective:

- What elements of U.S. force posture have caused allies to feel assured or alternatively to develop their own deterrent capabilities?

Actor Type

- How have states successfully and unsuccessfully deterred non-state actors from using violence?

Phase

- What strategies has the United States used to effectively de-escalate conflict after it has begun?

Threat

- Under what conditions have states agreed to abandon their nuclear weapons development programs?

International Future

- Does nuclear proliferation tend to occur when great power relations are cooperative or competitive?

Case study research can contribute to mid-level theory, which can then inform broader questions about effective strategies and force postures.

Case studies are particularly important for determining the scope and generalizability of particular events or phenomena. Given that nuclear events, including crises, proliferation, and use are rare, analysts are often left to reason by analogy. For example, lessons are commonly drawn from the Cuban Missile Crisis or, more generally, from the Cold War experience. While these events provide important sources for data and theory building, comparison must be made in structured ways to ensure they are done in a methodologically sound manner. This is the role of structured case studies.

Case studies are also extremely important for focusing on what are sometimes called "deviant" case studies. These are important special cases that populate strategically significant corners of the possibility space. For instance, even if large-N quantitative analyses identify broad principles for strategic approaches across a broad range of actors, it may be the outliers in those models that deserve examination. To be concrete, a U.S. strategy that effectively deters every country in the world except Iran and North Korea should not be deemed sufficient. In this instance, it is the specific outliers, not the general phenomenon, which should serve as the focus of research.

FURTHER RESOURCES

There is a vast literature on case study and comparative methodologies. This review relied heavily on: George, Alexander and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.

For a canonical example of the comparative method on the topic of nuclear weapons and deterrence, see: George, Alexander and Richard Smoke. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. New York: Columbia University Press.

Other Leading Works Include:

Bennett, Andrew and Colin Elman. 2007. "Case Study Methods in the International Relations Subfield." *Comparative Political Studies* 40(2): 170-195.

Collier, David. 1991. "The Comparative Method: Two Decades of Changes." In Dankwart Rustow and Kenneth Paul, eds. *Comparative Political Dynamics: Global Research Perspectives*. New York: Harper Collins.

Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In Fred Greenstein and Nelson Polsby, eds. *Handbook of Political Science*, vol. 1, *Political Science: Scope and Theory*. Reading, MA: Addison-Wesley.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Lijphardt, Arend. 1971. "Comparative Politics and Comparative Method," *American Political Science Review* 65, no. 3 (September): 682-98.

Mahoney, James. 2000. "Strategies of Causal Inference in Small-N Analysis" *Sociological Methods & Research* 28(4), May 2000: 387-424.

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley and Los Angeles: University of California Press.

Trachtenberg, Marc. 2006. *The Craft of International History: A Guide to Method*. Princeton: Princeton University Press.

SUBJECT MATTER EXPERT ELICITATION

Identify the levels, data observations and forms for which this method is most suitable.

Level

- ❑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ❑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Time-series**: multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ☑ **Qualitative**
- ❑ **Quantitative and qualitative**

INTRODUCTION

Changes in nuclear posture and policy occur in the international stage in a dynamic environment. SME elicitation enables the researcher to discover insights into the far-reaching implications of changes to force posture and policy, especially how these may affect international relationships. It also provides a longer timeframe by examining not only the initial responses of actors, but analysis of extended changes to the status quo through several iterations of responses and counter measures adopted by pertinent actors.

SME elicitation allows analysts to collect multiple perspectives that, together, produce a more robust picture of the complex research issues. Standard analytic techniques, such as content or narrative analysis, can be used to sift through the data to discover key insights or themes. Equipped with these insights, decision-makers can make more informed choices about force structure and strategy in the dynamic nuclear environment.

METHODOLOGICAL OVERVIEW

A subject matter expert (SME) is an individual who, by virtue of position, education, training, or experience, is expected to have greater than normal expertise or insight relative to a particular discipline (Pace, 2002). SMEs are typically at the forefront of a specialty relevant to the problem and are recognized by their peers as authorities because of their sustained and significant research on the topic (Kotra et al, 1996).

SME elicitation is a formal, highly structured, and well-documented process in which judgments, usually of multiple experts, are obtained. Typically, an elicitation is conducted to evaluate an area of uncertainty due to insufficient data (Korta et al, 1996). An SME elicitation procedure should be developed to minimize inherent biases in subjective judgment and errors in the elicited outcomes (Slottje, 2008).

Eliciting information can be considered more of an art than a science. The success of SME elicitation depends on intangible factors, such as the rapport between the interviewer and SME and their individual communication skills. SME elicitation, as employed in the CANS effort, was shaped by the guiding principle of preserving multiple points of view and exploring differences, rather than insisting on achieving a consensus. The CANS team elicited SME information through the use of interviews, models and simulations, war games, crowdsourcing, and academic outreach to ultimately assess nuclear weapons policy objectives, threats, and international environments.

HOW IS THIS DONE?

The SMEs selected for elicitation should be individuals who: (a) possess the necessary knowledge and expertise; (b) have demonstrated their ability to apply their knowledge and expertise; (c) represent a broad diversity of independent opinion and approaches for addressing the topic in question; (d) are willing to be identified publicly with their judgments; and (e) are willing to publicly disclose all potential conflicts of interest (Kotra et al, 1996).

The specific processes and tools for elicitation vary by individual project requirements, but the basic steps can be summarized as follows:

- Identify the data that is needed and potential sources
- Develop data collection strategy and tools
- Conduct SME elicitation
- Follow up with sources as needed

The CANS project utilized interviews, models and simulations, games, and other sources (crowdsourcing and academic outreach) to effectively elicit SME information.

INTERVIEWS

Interviews can be an effective method of eliciting SME information. There are three main types of interviews that can be used for SME elicitation: open-ended, structured, and discussion groups.

Open Ended

Open-ended interviews feature an interviewer having a conversation with an SME without having a set list of questions. The benefits of open-ended interviews are that they are especially useful for collecting qualitative data, are helpful for capturing a first-hand narrative of events, and can build rapport with SME that can ease concerns and improve outcomes. The drawbacks of open-ended interviews are that the SME may not remember quantitative data accurately, qualitative data must be coded for statistical analysis, and poorly worded questions can result in response bias.

Structured

Structured interviews feature an interviewer following a standard set of questions with each SME. The interviewer may pursue additional lines of inquiry with the SME if research design allows. The benefits of structured interviews are that the interviewer can elicit qualitative and quantitative data, it ensures the interviewer covers all relevant topics, and it increases consistency when there are multiple interviews. The drawbacks of structured interviews are the interviewer may miss opportunity to probe the SME for alternative beliefs and perspectives and that poorly worded questions can result in response bias.

Discussion Groups

Discussion groups feature several SMEs discussing varied experiences to explore an issue or topic collaboratively. The discussion groups can participate in divergent or convergent thinking activities. Divergent thinking activities develop new ideas and identify alternative perspectives, while convergent thinking activities develop solutions to problems and identify overall themes from the discussion. Discussion groups are beneficial because they develop a multi-faceted view of a topic, can be helpful for obtaining SME validation of research questions or conclusions, allow SMEs to challenge each other to clarify the groups' collective views, and can observe how SMEs views evolve. The drawback of discussion groups is that the discussion must be managed by a facilitator to ensure all views are heard, de-escalate strong emotional reactions and debates, and ensure the discussion remains focused on objectives.

GAMES

Games bring SMEs, typically from a variety of backgrounds, together to play a specific role within a given scenario. Teams are developed to discuss the related issues and create collective views according to game objectives. Games are beneficial to observe how SME opinions evolve through the course of the discussion and to examine intended and unintended effects of new strategy or policy. The drawbacks to games are that success is heavily dependent upon the facilitator's ability to manage discussion flow and the design, development, implementation, and evaluation of the game can be labor intensive. Additionally, it is critical that the game organizers make sure that the right SMEs are playing the right characters in the game in order to increase accuracy and reduce bias.

MODELS AND SIMULATIONS

Models and simulations use independent mathematical models or a suite of interacting models that include a number of variables. Researchers can input a range of values for each variable over a large number of trials to understand the impact of each variable and the interaction between variables. The benefits of using models and simulations are that they combine proven analytic theories and methods from various fields and can show the effects of decisions in a multivariate environment. The drawbacks of models and simulations are that the models' validity may be questioned, and the designing, development, implementation, and evaluation of the models and simulation can be labor intensive.

CROWDSOURCING

Crowdsourcing is an advanced analytic technique that involves tapping a diverse set of SMEs to develop a multi-faceted perspective of the research question. Crowdsourcing is beneficial because independent SME inputs contribute to data reliability and source validation. This technique also allows for multiple SMEs to be tapped in hard-to-reach conflict environments allowing for on-the-ground insight into a specific problem. The drawbacks of crowdsourcing are that reaching qualified SMEs may be difficult for certain topics and regions, and validating SME identity may be difficult if using online methods.

ACADEMIC OUTREACH

Academic outreach involves asking scholars from a variety of fields and institutions to develop high-quality papers on given topics or issues that are suitable for publication in a peer-reviewed journal. Academic outreach is beneficial because it combines analytic methods to produce a rich understanding and can include both established and emerging scholars. The drawbacks of an academic outreach are that the existing body of research may be insufficient to support any conclusions.

REQUIREMENTS

DATA

Low (2)

The type of data collected for this method is typically qualitative, but this method can be used to validate quantitative data. SME elicitation is context-dependent, as it will the identification of appropriate SMEs. Existing quantitative data sets may exist for some fields and may be used as a starting point for SME elicitation interviews.

SET UP TIME

Short (2)

An effective SME elicitation requires an understanding of the information that is needed or missing, an idea of which SMEs can provide the needed information, a well-developed data collection strategy and tools for data collection, an effective elicitation of information from the SMEs, and a review and follow-up period of the information obtained. The time required for this analysis can vary depending on the scope and magnitude of the questions being asked, as well as the specific SME elicitation methods being used. Selecting appropriate SMEs and developing the specific methods to be used for SME elicitation can vary; developing a game or model and simulation will require more time than developing a structured questionnaire for an interview.

EXECUTION TIME

Long (1)

Once the questions have been developed and the SME selection process complete, the time required to elicit the information is fairly short. Factors that can influence the execution time include SME availability, level of detail required for data and whether information can be collected through a mail (or email) survey or requires interviews with the researcher.

SKILL SET/EXPERTISE

Specialized-advanced (2)

Access to a diverse network of SMEs is necessary for an effective SME elicitation. Different methods of elicitation, whether it is interviews, models and simulations, games, crowdsourcing, or academic outreach, have varying requirements for the type and amount of SMEs. It is crucial to have access to SMEs in the field of the area of uncertainty for the specific question. It is also beneficial to elicit information from multiple SMEs for any given problem. Familiarity with content analysis and other analytic techniques will help the researcher make sense of the interview data and identify key themes.

TOOLS

Minimal (0)

No specific tools are required for this technique, however the researcher does need to have 1) and understanding of the research topic (to have an idea of what perspectives are needed and likely points of contention); 2) a method for accurately noting SME responses (take notes, recording the interview); 3) access to SMEs, and 4) a complete understanding of SME elicitation techniques

COST

$ (7)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

SME elicitation is an effective method for addressing all of the general questions identified in the 5D framework. The key is to identify a set of SMEs such that the researcher hears multiple perspectives on force posture and policy.

Considerations and recommendations about optimal strategies and optimal force posture are context dependent, and SME elicitation results are not generalizable beyond the specific situation assessed. However, broader themes about policy or force structure issues may be useful as starting points for future SME elicitation activities.

SME elicitation relies upon human memory, and as such, it may not be the most effective way to discover specific details, such as dates or times. Researchers can combine methods (e.g., use open source research to find records) to develop a log of events and use SME elicitation to validate the events and ask qualitative questions, such as the meaning of these events or opinions.

FURTHER RESOURCES

Kotra, J. P., Lee, M. P., Eisenberg, N. A. DeWispelare, A. R. (1996). *Branch technical position on the use of expert elicitation in the high-level radioactive waste program.* Division of Waste Management: Office of Nuclear Material Safety and Safeguards: U.S. Nuclear Regulatory Commission.

Pace, D. K., Sheehan, J. (2002). *Subject matter expert (sme)/peer use in m&s v&v.* John Hopkins University and Defense Modeling and Simulations Office.

Slottje, P., Slujis, J. P. van der, Knol, A. B. (2008). *Expert elicitation: Methodological suggestions for its use in environmental health impact assessments.* National Institute for Public Health and the Environment Centre for Environmental Health Research.

SURVEY RESEARCH

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ❑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ☑ **Time-series:** Multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ❑ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

Surveying is a data collection methodology that queries representatives of a group of interest in order to generalize about the needs, attitudes, beliefs, behaviors, etc. of the group. Often surveys are used in the social sciences to understand a particular group's behavior, perceptions, and attitudes. People turn to surveys when they have a pressing information need and current information sources are insufficient, inappropriate, or nonexistent. One of the greatest benefits of surveys is that a relatively small group of the overall population can inform a researcher on that population's knowledge, attitudes, beliefs, expectations, and perceptions with relatively few resources, expenditures, and effort, if that group is representative of the population of interest[29].

In the United States, surveys are often used in advance of elections to anticipate election results or to identify issues of major concern to voters. Additionally, surveys are used extensively by the United States Government (USG) to estimate unemployment rates, supplement the census, and estimate land use.

Surveys can be used iteratively to identify trends or can be used as a snapshot in time. While surveys can be conducted with relatively minor technical expertise, they require careful planning, an investment of time and, depending on the methodology employed, significant financial and personnel resources.

[29] A survey differs from a census in that only a representative sample of the population is studied, not the entire population.

METHODOLOGICAL OVERVIEW

A survey is data collection methodology that gathers information from a sample of a population of interest. A **sample** is a smaller, representative slice of the population of interest. For example, if you were interested in understanding the effect of changes to Medicare prescription drug coverage, the population of interest would be Medicare recipients. The sample must be representative of the population in order to draw accurate inferences about the effects of the phenomenon (in this case, changes to Medicare), on the population as a whole. A **representative sample** is drawn "from a population so that particular properties of the population can be estimated accurately from the sample. For example, political scientists may draw samples from the population of voters to predict with some certainty the outcome of an election" (Statsoft, n.d.).

Surveys can be classified by:

- Size (large-n, straw poll)
- Type (telephone, mail, web, in-person interviews)
- Content (opinions, attitudes, behaviors, factual information)

Table 2. Benefits and Limitations of Survey Types

	Mail	*Telephone*	*Web*	*In Person Interviews*
Benefits	Least expensive Best for sensitive questions No interviewer bias	Inexpensive Fast Can support long questionnaires Supports open ended responses Moderate cost	Least expensive Can be administered to large populations Supports complex survey designs Good for unscientific straw polls	Best for complex or open ended questions Highest response rate Lowest burden on respondent Often used to collect information on attitudes, perceptions and opinions
Limitations	Low response rate Poor choice for open ended questions Slower than phone or online surveys	Low response rate Those with listed telephone numbers may not be representative of population of interest	Almost impossible to assure random sample Not great for long questionnaires	Expensive Possibility of interviewer bias Respondent bias on sensitive items Slowest form of data collection High cost

HOW IS THIS DONE?

There are eight primary steps common to any mode of survey design. These are listed below. A survey that is conducted to shed light on complex social science issues, such as political durability or political violence, should be designed to fit into a survey model. A survey model is "a social science model of any phenomenon or outcome that can be depicted graphically as a set of variables (represented by circles) connected by a set of causal effects (represented by single-headed arrows) and correlations (represented by double-headed arrows)" (NCSU, n.d.). This could also be referred to as a conceptual model. When dealing with complex social issues, the questionnaire should be designed to relate to variables in the model to help improve understanding of a social science phenomenon.

For example, a researcher may want to better understand political durability. If the researcher hypothesizes that a strong sense of nationality is an indicator of a country's political durability, he might consider conducting a survey to determine the degree of national identification in a country. To get to this issue in a given country (e.g., Pakistan), he may form a survey question, "What do you consider yourself first: a Pakistani or a Muslim?"[30]

1. **Choose the sample:**

The sample should be reflective of the population of interest. To determine this, the design team should create a sampling frame to identify the parameters of the population the survey is designed to research. A **sampling frame** is a list of all those in a population of interest who can be sampled. Depending on the research design, it may include individuals, households, or organizations. Population parameters can be defined by geographic boundaries (like a zip code) or by authoritative documents, such as an electoral register or phone book. The quality of the sampling frame is primarily determined by the accuracy of the population's parameters. Several common problems can undermine the accuracy of a sample frame. Sample bias is one potential problem. For example, 5% of households in America do not have phones. These households are often low-income households. Therefore, a sample taken exclusively from the phonebook may introduce bias. Other problems abound, such as the use of caller ID to screen calls or the timing of the calls during working hours when many people are not home. Furthermore, it is essential that the sample has been randomly selected from the population so that "all of the units in the population have a known, positive chance of being selected" (Scheuren, n.d.). This can be difficult in the case of amorphous populations, like homeless men. Adequate description of the sampling frame allows for a calculation of sampling error. This helps determine whether the results are scientifically valid enough to draw inferences from the sample about the population.

[30] Pew Global Attitudes Project. (2011, July 21). Common concerns about Islamic extremism: Muslim-Western tensions persist. *Pew Research Center*. Retrieved from http://www.pewglobal.org/2011/07/21/muslim-western-tensions-persist/

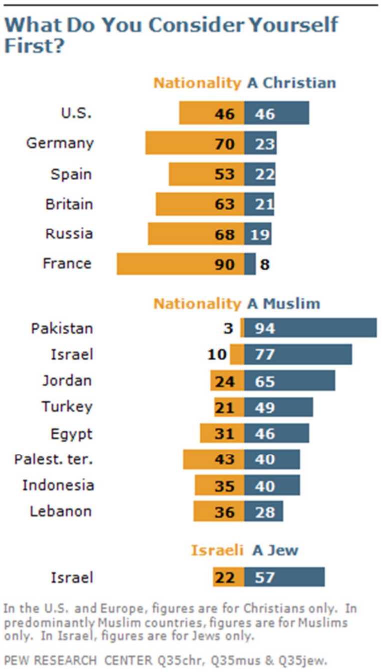


Figure 6. Pew Global Attitude Project, Muslim-Western Tensions Persist

Using the example of determining political durability through the self-identification of national identity, the population size is large—an entire nation of people. The researcher limits the population of interest to Muslim Pakistanis since he wants to know whether the respondents identify primarily with their religion or their nationality first. In the Pew survey, the sample size in Pakistan was 1,251 individuals (Pew, n.d.)

One common error in sample selection is to try to accurately match major demographic characteristics (age, gender, race, etc.) in the sample to the population (referred to as quota sampling). Demographic characteristics are only important to include in the representative sample if the characteristics are expected or known to be related to the outcome being measured. It is, therefore, sometimes better to draw a random sample of the population (referred to as probability sampling) when it is unknown whether characteristics of the population are related to the outcome. Finally, the size of the sample required is determined by the size of the population and the statistical quality (i.e., margin of error) desired.

2. Plan survey instrumentation

Survey instrumentation is the step where the questionnaire is developed. This step is often considered the most critical phase where the design team (1) determines the mode of data collection (mail, telephone, web, in person) and (2) clearly defines the concepts and questions. Failure to clearly define the objective of the survey and to carefully select, phrase, and order questions could result in seriously misleading results.

The survey design team needs to be particularly careful to avoid survey bias during the questionnaire design process. The manner in which a question is asked can greatly affect the results of a survey. For example, a recent NBC/Wall Street Journal poll asked two very similar questions with very different results: (1) Do you favor cutting programs such as social security, Medicare, Medicaid, and farm subsidies to reduce the budget deficit? The results: 23% favor; 66% oppose; 11% no opinion. (2) Do you favor cutting government entitlements to reduce the budget deficit? The results: 61% favor; 25% oppose; 14% no opinion (Ferber, 1985). Because people may not understand that government entitlements in question 2 refers to social security, Medicare, Medicare, and farm subsidies, they may be more likely to want to cut the programs compared to when the entitlements are spelled out in question 1. Politicians may rely on clever wording of polls to show evidence of support for their policy position, but good social science that seeks accurate results will attempt to eliminate bias.

Bias is often introduced in the way questions are worded. One academic in the health sciences listed 48 ways bias can be introduced in questionnaires (Choi & Pak, 2005). Even though the document is focused on the health sciences, it is entirely applicable to complex social science issues. Some common mistakes that result in bias include ambiguous questions, which allow the respondents to interpret the question in a variety of ways. For example, "Is your work made more difficult because you are expecting a baby?" Another mistake is the use of jargon, which many not be understood by the population of interest. For example, "What was your age at menarche?" Some questions introduce bias because they ask a sensitive question that makes the respondent uncomfortable. For example, "What is your religious affiliation?" In another example, if the questionnaire is too long, people may tend to respond with all "yes's" or all "no's." On the other hand, open-ended questions pose problems because the quality of response may vary greatly, since some people are not willing to invest time into a carefully thought-out answer. Another common source of bias occurs when respondents try to make the situation out to be worse than it is in order to obtain more resources. For example, "Do you have enough wells in your village to provide an adequate source of water?" Answering "no" may lead to the greater resources being allocated to that village. This is why Step 4 (pre-testing the questionnaire) is so important.

It is also important to determine whether you want your questions to yield qualitative or quantitative results. While most survey responses can be coded to be analyzed using quantitative/statistical analysis, responses can vary along a spectrum of responses. Variables in survey methods are referred to as nominal, ordinal, or interval/ratio. **Nominal variables** are categorical; they are groups that cannot be ranked or ordered. Nominal variables include yes/no, gender, ethnicity, etc. questions. **Ordinal variables** can be ordered or ranked and are usually represented using a scale (poor, fair, good, excellent). **Interval/ration variables** are measurable and numeric such as (age, income, temperature, and percentages). It is easier to process interval data, since it is already in a numeric format, but nominal and ordinal variables can be coded in order to make the data accessible by statistical analysis software.

The Pew survey (n.d.), was conducted by face-to-face interviews with their sample. The researchers acknowledge pulling the sample from a disproportionately urban population, but data are weighted to reflect the actual urban/rural distribution. They determined that the sample covered 85% of the population of Pakistan. They selected nominal variables: respondents could say they are primarily Muslim or primarily Pakistani.

3. Train the interviewers & supervise the interviews

The survey is only as good as the people conducting the survey, particularly for telephone and in-person interviews. Interviewers must be trained to encourage people to participate in the survey and to conduct the interview in a consistent, scientific manner. Furthermore, the conduct of the surveys must be supervised and reviewed at several points during the data collection phase to ensure quality and accuracy of data.

4. Pretest questionnaire and field procedures

The only way to make sure the questionnaire solicits the information it is designed to, is to conduct a pretest of the questionnaire and field procedures. The pretest can identify unanticipated bias and ambiguities that are sometimes difficult to anticipate during the survey creation stage. See the discussion on sources of bias in step 2.

5. Collect data

Data collection methodology varies considerably according to the mode of survey employed. Regardless, it involves a process of asking people questions and recording their responses. Data collectors (for phone and in-person surveys) have to be particularly careful not to inject bias into the data collection process, but web and mail require less effort to collect data and run the least risk of introducing response bias. The disadvantage with these less personal collection methods is a decreased response rate and the lack of opportunity to clarify any uncertainties the respondent might have about a survey question. Person-to-person data collection methods often result in more bias due to the potential for the interviewer to infer questions, to provide leading information or explanation about the question or survey, to misunderstand the respondent, or to insert their own bias in the responses.

6. Process data

Processing data is often labor intensive and could require hours of coding to standardize responses to open-ended questions. Data is more easy to process if the questions are categorical (male/female), ordinal (low, medium, high), or interval ($10,000, $100,000). They are also easier to process if the respondent is asked to select a response based on a likert scale (strongly agree, agree, neutral, disagree, and strongly disagree). Web-based surveys offer a distinct advantage in data processing, as all data are collected digitally and require minimal coding and data manipulation prior to analysis.

In the Pew example, the respondents were asked a categorical question: what do you consider yourself first—a Muslim or a Pakistani? Processing and coding the responses is relatively easy for a categorical question like this.

7. Analyze the data

The standardized data obtained from the survey is usually analyzed using statistical software such as SPSS, SAS, or R. However, survey data that is qualitative in nature (e.g., responses from open-ended questions) is usually processed through descriptive or content analysis. This method requires the analyst to read the survey results and identify trends or themes and how they relate to one another. Alternatively, a coding scheme could be used to transform the qualitative responses into quantitative data.

An example of an open-ended question is, “what are the political issues that are most important to you?” The open-ended question allows the respondent to accurately describe their most important concern without being limited to a pre-defined set of responses. However, the open-ended responses must be broken down and categorized in some methodical fashion. Results can be manually coded and grouped into logical responses (e.g., abortion, economy, the war, etc.). Alternatively, results can be manually coded by content analysis software or natural language processors.

8. **Report the survey's findings**

A critical element of reporting is the margin of error. A margin of error is often represented as follows: "55 percent of respondents favor Ms. Smith in the upcoming mayoral election. There is a margin of error of 3 percentage points." The **margin of error** is a common summary of sampling error that quantifies uncertainty about a survey result. Sampling error refers to factors as faults in sampling, coding, tabulating, data processing, interviewer bias, researcher bias, and data misinterpretation. Larger samples tend to have lower margin of error because the high number of respondents is more likely to reflect the overall population. The margin of error is primarily affected by the sample size, the type of sampling done, and the size of the population.

Sampling error could also result when the margin of results is very narrow. For example, if candidate A is preferred by 52% of the respondents compared to Candidate B, who is preferred by 48% of the respondents, and the margin of error is 2%, it would be unfounded to claim candidate A is in the lead.

Survey Quality and Sources of Bias

The quality of the data collected from surveys is high dependent on accuracy of the sample, which is a function not only of the sample frame design, but also the response rate (percent of respondents who fill out or answer questionnaire). It is better to have a small, representative sample than a larger sample with a response rate lower than 50%. If the response rate is low, then the representativeness of the sample is compromised, and results cannot be extrapolated to the population. Poor survey results can result in bias (as can poorly worded questions—see step 2), which inhibits the survey team from making conclusions about the populations based on the sample. Sources of bias include the following (NCSU, n.d.):

- *ambiguity of questions*: Questions should be specific, avoiding generalities. For example, on a scale from 1 to 10, how popular is President Clinton at this time? This example begs the question, popular with whom?

- *lack of mutual exclusivity when multiple responses allowed:* When multiple response items are allowed, bias is introduced if the items are not mutually exclusive, yet only a single item may be selected.

- *non-exhaustive response set:* Bias is introduced when the response alternatives available to the respondent leave out valid choices they would otherwise make. The most common example is leaving out such responses as "neutral" or "don't know" when, in fact, respondents may well be neutral or may actually not know, rather than be hiding their "true" responses, which the researcher is trying to force out by omitting these categories.

- *residual categories (don't know, not applicable):* "Don't know," "Don't care," "Not applicable," and "Did not respond" are separate types of responses. It is better for the instrument to keep them separate so they can be analyzed for patterns, even if they are combined for other purposes in statistical processing.

- *rank lists:* Ranking can be a challenging task. Many survey researchers recommend that respondents not be asked to rank more than four or five items. Beyond that, respondents may give arbitrary rankings just to get past the item.

- *social favorability:* Certain topics deal with actions or beliefs for which there is a socially favored position (people are supposed to enjoy parenting, to vote, not to approve of marijuana use, etc.). Such items should be buffered with phrasing such as "Some people favor x, some people favor y, which do you favor?"

- *loaded terms:* Ex. "Do you lean more toward the pro-life or toward the pro-abortion position on issue of termination of late-term pregnancies where the health of the mother is threatened?" This example is biased because one position is labeled with its most favorable label (pro-life, rather than anti-abortion), while the other position is labeled with its less favorable label (pro-abortion, rather than pro-choice).

- *leading questions:* "Do you favor an increase in the federal minimum wage to $8.00?" is slightly leading because it does not legitimize both affirmative and negative responses.

- *unfamiliar terms and jargon:* Ex., "Do you consider environmental regulation of wetlands to be an invasion of the sovereignty of the states?" Terms such as "sovereignty" are apt to not be well understood by typical survey populations. Wherever possible, familiar terms should be substituted for unfamiliar terms.

- *requiring inaccessible information:* An item may use familiar terms but require information most respondents would not know. For instance, a question about Stalin's treatment of the Cossacks might have been acceptable long ago, but today's population of respondents is apt to know little about this subject and perhaps not even recognize "Stalin."

- *Multidimensionality*: A form of ambiguity arises when items are multidimensional. Ex.: "On a scale of 1 to 10, please rank the performance of the president?" The respondent may be torn between multiple dimensions: personal vs. official performance, or domestic vs. foreign policy performance, for instance.

- *compound items:* Items with compound clauses may not be multidimensional but may involve undue complexity (see below). For instance, the item, "Do you have or have you ever had a physical, mental, or other health condition which has lasted over six months and which has limited the kind of work you could do on your job?" is better broken into two items: "Do you have or have you ever had a physical, mental, or other health condition which has lasted over six months?" and the follow-up item, "If you answered yes to the previous question, did this condition limit the kind of work you could do on your job?"

- *recall items:* People's ability to recall the past is limited. The more current and specific the question reference, the better. If recall is necessary, the time frame should be as recent as possible and not over six months unless the reference is to major events (ex., marriage, changing jobs, buying a car).

- *complexity and memory overload:* It is possible to overtax the respondent by requiring an excessive memory burden due to complexity. The more complex the item, the easier it is to overload memory.

- *poor grammatical format:* Weak grammatical format can introduce bias. For instance, the item, "Would you say that you approve very strongly, strongly, ..." presents "dangling alternatives" which the respondent must memorize before even knowing what the question is. This format is frustrating and may bias responses toward the first-presented response or toward negativity. Putting the subject first is the preferred order.

- *hypothetical items:* Hypothetical items (ex., "What would you do if ...") creates a difficult challenge for respondents. Seriously considering such items requires time for imagination and consideration. People tend to base responses to such items on their most-related actual experiences, and it may be better to ask about such experiences directly.

- *inappropriate assumptions:* Items should not contain false or arguable premises. The respondent should not be required to make a false or arguable assumption in order to respond to an item on its face. For instance, "How much improvement in your life would passage of the Equal Rights Amendment make? A great deal, some, very little, or none?" is an item which assumes that the effect of ERA passage could not be negative on a respondent, forcing respondents who believe the effect would be negative to either skip the item or give a response which does not represent their views.

- *gender bias:* Avoid gender-specific wording in items. Not only may some respondents take offense, but also some respondents will interpret a reference to "man,"" for example, to refer to humans while others will interpret the reference to refer to males, thereby creating an unwanted response ambiguity.

- *mismatched item and response set:* The response categories should be appropriate to the dimension probed by the item. For instance, "How unusual do you think it is for a person to donate $100 or more to a presidential candidate? A great deal, some, very little, or none?" is a mismatch because "a great deal unusual," "some unusual," etc., are not grammatically acceptable responses and will confuse respondents.

- *language differences*: One must assure that items have the same meaning when the questionnaire is to be administered to populations speaking different languages. This is addressed by Behling and Law (2000). Simple direct translation of validated U. S. scale items into another language often will not create a valid scale in the other language. One must consider three dimensions of the problem: semantic equivalence, conceptual equivalence, and normative equivalence of items. An item may be acceptable in one dimension but not in another. Semantic equivalence is sought through the translation/back translation method, having independent translators translate from one language to another and then back again to see if the original and re-translated item remains the same.

REQUIREMENTS

DATA

Average-high (3.5)

The size of the survey is a function of two considerations: the quality of results desired and how the results will be used. In the United States, reputable surveys poll approximately 1,000 people in order to get a sufficient understanding of national opinions or attitudes. In general, the larger the sample size, the more confident one can be that the results truly reflect the population. Additionally, the larger the sample size, the smaller the margin of error. The data collected from surveys can be qualitative (Do you feel your political voice is represented by your elected officials? Why or why not?) or quantitative (How many people live in your household?). Likewise, the results of the survey could be qualitative (overall, the people of Helmand do not feel their elected officials actively represent their interests to the central government) or quantitative (the average household size is 2.59 people in the United States).

SET UP TIME

Short (2)

The shortest period for a simple survey using the telephone could be as little as 2-3 weeks. However, a national level survey with 1,000 respondents could take anywhere from three months to a year from initial planning phase through reporting results. While some steps can be done concurrently (defining sample frame and developing questionnaire), many of these steps are time consuming.

EXECUTION TIME

Long (1)

The length of the questionnaire does not always determine the time required to collect data. For example, the Tactical Conflict Assessment and Planning Framework (TCAPF) used by coalition forces in Afghanistan asks four simple questions to better understand the Afghan population's perspectives. However, the TCAPF survey is a time and resource intensive endeavor due to training requirements; the need to travel to remote, potentially hostile environments; and the need to spend time with the population to encourage participation.

A single survey cannot show trends or a change in attitude or opinion. Two or more surveys are required to establish patterns. In order to establish trends with a strong degree of confidence, the same survey questionnaire must be employed. Changes in attitudes can be identified between two surveys, but identification of trends requires at least three surveys.

SKILL SET/EXPERTISE

Specialized-average (1)

Professional survey researchers typically have educational backgrounds in social sciences, economics, mathematics, statistics, sampling theory, and survey design. A firm grounding in statistical methods is essential for properly analyzing survey results and setting up the survey design. The people conducting the interviews, however, do not require a specific educational background; however, they must undergo training on how to conduct interviews. Good interviewers are both engaging and detail oriented.

Conducting scientifically sound survey requires an experienced survey designer (or team), staff to carry out the survey (interviewers, preparing mailings, preparing web survey), statisticians to interpret the data, and analysts to apply the results of the survey to the complex social problem of interest.

TOOLS

Average (1)

The analysis portion of the survey process typically requires access to statistical software such as SPSS, SAS, STATA, or R.

COST

$$ (8.5)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Surveys can help provide contextual information about the population of interest. In order to have a successful deterrence strategy, the United States Government (USG) has to understand what makes people comply. One of the ways to find out the attitudes and beliefs of a population is through surveys. Deterrence strategies could be constructed around these findings to help determine optimal US force postures and strategies. Surveys help educate a nuclear defense analyst how a population might respond to certain threats or rewards. It is less able to directly inform what kind of force posture should be employed; rather, it hints at the reaction to a certain course of action.

FURTHER RESOURCES

American Statistical Association. (2011). Survey research methods section. Retrieved from http://www.amstat.org/sections/srms/

North Carolina State University. (not dated). Survey research. *College of Humanities and Social Sciences is among the largest colleges at NC State University*. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/survey.htm

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. *Forum: Qualitative Social Research.* Vol. 11, No. 2, Art. 11. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/viewArticle/1450/2946

Statsoft Electronic Statistics Textbook. (not dated). Glossary. Retrieved from http://www.statsoft.com/textbook/statistics-glossary/r/?button=0#representative%20sample

Scheuren, F. (not dated). What is a survey? Retrieved from http://www.whatisasurvey.info/

Choi, B. C. K., & Pak, A. W. P. (2005, January). A catalog of biases in questionnaires. *Preventing Chronic Disease.* Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1323316/

Pew Global Attitudes Project. (2011, July 21). Common concerns about Islamic extremism: Muslim-Western tensions persist. *Pew Research Center*. Retrieved from http://www.pewglobal.org/2011/07/21/muslim-western-tensions-persist/

Behling, Orlando and Kenneth S. Law (2000). *Translating questionnaires and other research instruments: Problems and solutions.* Thousand Oaks, CA: Sage Publications. Series: Quantitative Applications in the Social Sciences No. 133.

Raosoft. (n.d.). Sample Size Calculator. Retrieved from http://www.raosoft.com/samplesize.html.

DYNAMIC NETWORK ANALYSIS AND MODELING

Identify the levels, data observations, and forms for which this method is most suitable.

Level

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

  This information can be used as input, but the tools speak to an individual decision maker's behavior only as it relates to the behavior of others. Thus, it can provide insight on Kim Jong-Il if there is information on whom Kim Jong-Il interacts with.

Observations

- ☑ **Time-series:** multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ❑ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

Groups are difficult at times to understand and predict. Dynamic network analysis and modeling (DNA&M) enables the assessment of groups as complex socio-technical systems. Such groups might be platoons, terrorists groups, tribes, nation state leaderships, general city populations, or global alliances. Dynamic network analysis is an approach for assessing groups as a set of actors and the relationships between those actors that is particularly suited to information that describes who interacts with whom, when and where. It allows the analyst to answer questions such as:

- Who are the key actors?
- What are some of the key relationships in this group?
- Is this group cohesive or fractured?
- How resistant is the group to random and targeted change?
- Who are the likely future leaders of this group?

Dynamic network analysis techniques can provide answers to all of these questions. If the group in question has been studied for a sufficient period, these techniques can help identify the group's patterns over time. These questions and patterns include those below:

- Does the group/network have a regular cycle of interaction? (e.g., a meeting at the start of every month between team leads)
- Is the group becoming more or less cohesive?
- Do key actors cycle in and out of importance, or is there a stable leadership coterie?
- Key change detection: Is the current evolution of the network "normal" for the group?

Dynamic network analysis and modeling is an approach that assesses groups using a combination of statistical network analytic techniques, graph theoretic metrics, artificial intelligence and machine learning algorithms, visual analytics, and agent-based dynamic-network computer-based simulation. Dynamic network analysis and modeling identifies and illustrates relationships based on certain characteristics of group interactions such as who, what, how, and why through time and geo-location. Using DNA&M tools, the analysts can extract information about groups as networks; analyze these groups and identify key actors, sub-groups, patterns of collaboration and coordination, performance characteristics, regions of influence, capabilities; assess change in these groups through time; and identify the impact of various interventions or courses of action on the way in which these groups change in terms of their cohesion, performance, sentiment/beliefs, or activities.

Network analysis is richer and more useful if analysts represent more than merely people and their relationships in the models the analysts build. Where and in what context actors meet can say almost as much as how often they meet. If two people meet at an office, that tells you one thing; if they also meet up to play rounds of golf, that may tell you something else; and if they go to the same church or spend time at the same pub, that gives you even more insight into their relationship (and perhaps insight into the relevance of the pub, church, or golf course). Thus, although standard Social Network Analysis allows you to consider questions of "Who?," our approach allows an analyst to explore questions of "What?, How?, Where?, and Why?". These richer networks are what Carnegie Mellon University (CMU) calls **meta-networks**. Elements in meta-networks consist of any of the following: people, locations, resources, knowledge, events, tasks, roles, and beliefs.

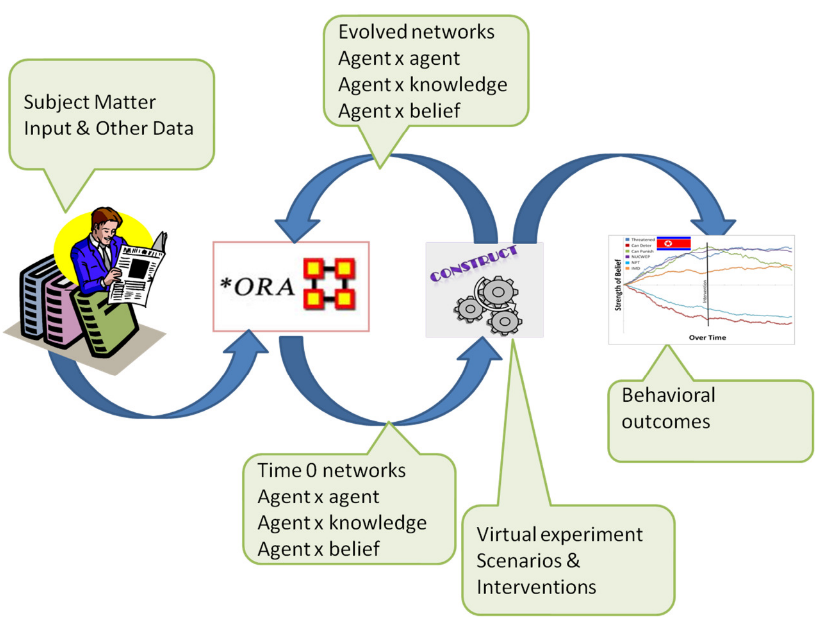
In addition to the set of network analysis tools included here, it can be worthwhile for analysts to be able to answer "What if…" questions. These questions require specialized tools to simulate the group's moment-to-moment dynamics, as well as the what-ifs of interest, which can include (but is not limited to):

- What if this actor was isolated from their group(s)?
- What if the people in this group(s) could no longer communicate?
- What if we inject rare knowledge into the group, how would the rare knowledge spread?

METHODOLOGICAL OVERVIEW

Dynamic network analysis and modeling is based on a suite of analytic techniques. These have been embedded into two core, easy-to-use technologies. These are Carnegie Mellon University's ORA and Construct models, which are available online. These technologies can be used together to reach important insights in how groups of interest behave, why they behave in those ways, and how that might change in different circumstances. ORA is a network analysis toolkit that can be used to enter, visualize, and analyze both social network (who talks to whom) and meta-network (networks data connecting the who, what, where, when, how, and why) information and about the group. Using ORA, the analyst can answer questions such as: "who is critical?, what are they doing?, how is it being done?, and where is their region of influence?" Construct is an agent-based dynamic network model that can be used to assess how groups and the related social/alliance network and knowledge network co-evolve under different intervention scenarios. Using Construct, possible futures resulting from changes in these networks can be explored through a series of "what if?" questions.

Once an analyst has created a meta-network (also called a model) of a group(s) of interest, s/he can use ORA to examine the network. The examination can determine, for example, important actors (e.g. powers-behind-the-throne) based on their centrality, as well as sub-groups within the network of interest. The analyst can then use Construct to explore the dynamics of the meta-network. The exploration can include examining how information is likely to flow through the group/network, given its current structure, as well as impacts of future events the analyst wishes to explore.



For DNA&M simulation, CMU uses agent-based dynamic network computational models (Construct) specifically engineered to support dynamic network organizational modeling and which were instantiated with dynamic network data created in, visualized by, and analyzed using ORA. The results were then assessed using dynamic network analysis tool (ORA).

Importantly, ORA can be used to create the "world" you are simulating. This is the instantiation or "time 0" input for the agent-based dynamic network model, Construct. Construct takes this "time 0" world and then simulates the behavior of the group by letting the agents interact and exchange information for some numbers of time steps, resulting in a new or "evolved" world. This information serves as guidance about how the networks are likely to change, the behavior and sentiments of the individuals are likely to change, and the overall changes in the group's activities. The output from Construct includes the evolved networks. These evolved networks can then be analyzed with ORA, and ORA is used to compare the "time 0" with the evolved situation to show the impacts of the interventions. Using ORA and Construct together in this way means that the analyst creates a meta-network of a group, uses ORA to assess the meta-network, and then uses Construct to simulate the impact of the interventions on these networks. The ORA/Construct tool combination is flexible, scalable, and applicable to a large number of domains. The models can be instantiated using quantitative or qualitative data, and the simulation module is reusable, both in the same situation to evaluate other scenarios and interventions and in similar situations. The simulation set-up and configuration can be set up so they are reusable with few modifications, allowing analysts to more easily examine other outcomes, apply the simulation and analysis to a new network/group of interest, as well as experiment with other possible inputs.

DYNAMIC NETWORK ORGANIZATIONAL MODELING

Agent Based Models (ABMs) can be used to simulate a group or organization. Exactly what is simulated and what data is generated varies from model to model, but behavior, information flow, process flow, and task execution can all be readily modeled. The "agents" in ABMs are called such because they have agency, the ability to affect both themselves and others through their actions. Frequently, we think of people or groups of people, or social media as being entities with agency (i.e., as agents).

In ABMs, the virtual world topology is typically a 2D grid-like quad paper, and agents form "networks", such as social networks of who interacts with whom, as they occupy the same or neighboring spaces in the grid. However, there is a special class of ABMs designed to be more realistic in terms of the social topology, the dynamic network organizational models. In dynamic network organizational modeling, the network is distinct from the spatial environment. The social network connecting agents is based on various socio-demographic, historical, and technological considerations. Ties between agents exist because the agents have something in common, and not just because they happen to be in neighboring spaces in a grid. Thus, rather than physical adjacency, a combination of social and physical adjacency is used to define the interaction spheres and networks. This social network topology may be static or dynamic. Moreover, within this social network, the agents are linked via multiple sub-networks; e.g., the formal authority and informal friendship network, or the alliance and adversarial network. Organizational dynamics, such as hiring, firing, training, and movement of personnel, all play into the interaction logics and impact who uses what.

These dynamic network organizational models have been used to assess changes within and among organizational units at both a single- and multi-level. Illustrative applications include impact of learning on organizational performance, merger assessment, leadership assessment, group performance, evolution of inter-organizational activity, assessment of terror groups, and identification of effective intervention strategies for counter groups associated with terror activities, narcotic trafficking, and insurgencies.

Using Construct for Network Organizational Modeling

In this example, the specific agent-based computational model used for dynamic organizational modeling is the Construct model. In Construct, the agents occupy a social network position defined in terms of which other agents they can interact with. Agents typically represent people, organizations, social media, artificial agents, countries; i.e., any type of information processing agent. This can be set by historical empirical data about who agents have interacted with or are currently interacting with. Construct operates at a middle level in terms of the cognitive realism of the agents and a high level in terms of the social realism of the agents. Key features of Construct are as follows

- Established sub-modules for various communication media, including cyber media;
- Multiple interaction logics based on fundamental well-validated social principles of hemophilic-based interaction, expertise search-based interaction, and co-work/collaboration interaction;
- Ability to be instantiated from real data at a qualitative or quantitative level; and
- Realistic inadvertent and intentional error models for the agents.

Construct has been used for a large number of scientific studies. For example, it has been used to examine covert networks, the impact of isolating leaders in al-Qaida and Hamas, the design of naval ship crews, mergers in corporate America, and communication of time-critical information to people in different US cities.

DYNAMIC NETWORK ANALYSIS

Dynamic Network Analysis is the study of dynamic meta-networks, i.e., the study of relations within and between networks connecting who, what, how, and why through time and space. The data is multi-modal, multi-link, and multi-level. The links and nodes can be attributed, and those attributes and indeed the values of the links, can change with time and even be probabilistic. Where social network analysis considers only the linkage of who to whom, dynamic network analysis moves beyond to put these social interactions into trails of who was where when doing what. Then, using a variety of research-supported metrics and techniques, the analyst can identify key actors, locations, and information of interest, identify emergent leaders, assess group structures, evaluate performance, locate hidden groups and patterns, assess change, and conduct vulnerability assessments. These techniques include graph-theoretic metrics, non-parametric statistical procedures for network data, machine-learning techniques, change detection techniques, and visual analytics.

A key feature of this approach is that groups are modeled as a set of networks. Quantitative and qualitative data can be used to define the relations. Then an assessment of the existing structure is done. If the data is time variant, change detection and temporal analysis are done. Then the data can be used to instantiate an ABM, specifically a dynamic network organizational model. With the ABM, model analysts conduct a virtual experiment to assess the impact of changes on these networks.

Areas where dynamic network analysis can be applied are varied and nearly limitless. Gaining an understanding of the structure of Al Qaida, assessing the ecological map of a food web, identifying hidden ports using AIS data, locating hidden groups within the Enron corporation; target identification for IEDs, and so on are all areas analysts have addressed using this approach.

Using ORA for Network Analysis

In this example, the ORA dynamic network analysis tool is used to conduct analysis of the over-time organizational model created by Construct. ORA is a dynamic network analysis toolkit that can assess social and communication activity and, indeed, any meta-network data. The structure of a social system is defined by the relationship among its personnel, knowledge, resources, tasks, activities, and belief entities (see also Appendix 3). These entities and relationships are represented by a dynamic meta-network. ORA contains over 100 metrics, which are categorized by their use to the analyst and, so, the question they address.

Metrics are also organized by input requirements and by output. ORA generates formatted reports viewable on screen or in log files and reads and writes networks in multiple data formats to be interoperable with existing network analysis packages. In addition, it is a suite of visual analytic tools, including 2D and 3D graph visualizations, NODEL, and various chart features for metrics. Figure 1 provides examples of ORA visual analytics.

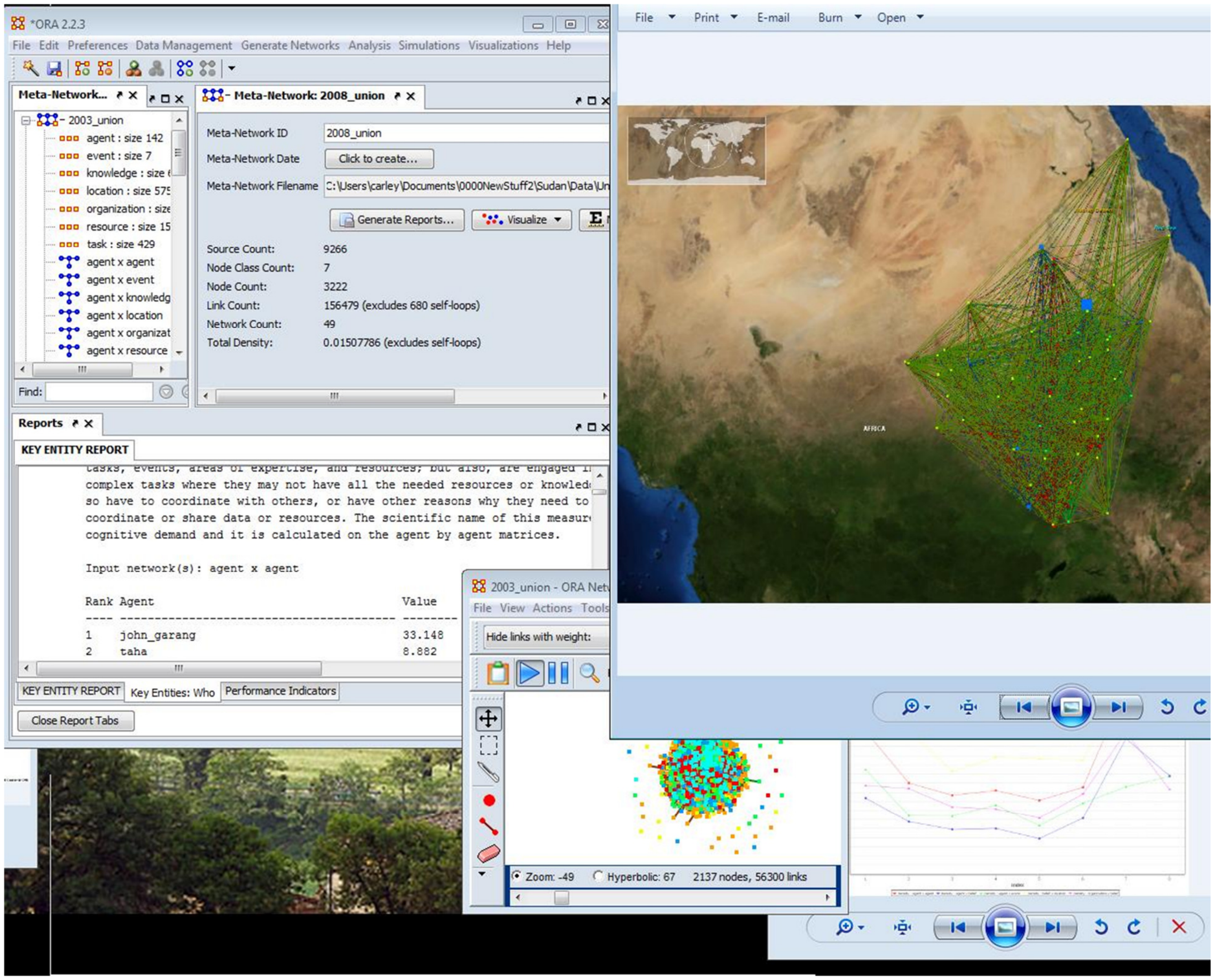


Figure 1. Illustration of typical ORA visualizations and analytics.

HOW IS THIS DONE?

INITIAL STEPS

Define the Question of Interest and the Hypothesis

The most basic steps for the experimenter(s) are to decide what question(s) of interest they are going to explore and what hypothesis(es) the analyst will attempt to validate or invalidate. Without these steps, the subsequent requirements elicitation and model generation can easily go out-of-scope.

Determine Resource(s) Availability and Constraints

After confirming with the client/customer that the analyst has accurately captured the questions to ask and answer, the analyst must begin by assessing the resources available to effort (e.g. time, computing cycles; data storage; subject matter experts; data sources). Further, an analyst must identify any constraints placed on him/her by the client/customer, as well as identify any specified and implied tasks that accompany the primary question of interest. With this information, the analyst is now in a position to develop a backwards-planning timeline that will help develop and sustain a schedule of completion for the overall effort.

Build the Model, Make It Better Over Time

After initial elicitation and planning is well on its way to completion, the analyst has enough information to develop and build the model. Decisions about which real-world entities will become abstracted entities in the Construct model, as well as the level of abstraction, help maintain model conciseness and clarity (e.g. model an entire country as a single monolithic entity or a collection of entities that are relevant to the question of interest). The analyst must be prepared to increase the level of abstraction (e.g. zoom out from the starting point), as well as decrease the abstraction (e.g. zoom in) to overcome ambiguous outputs in the experimental environment. Creating the nodes can be as simple, though potentially tedious, as manual insertion of nodes of various types on a drawing palette (e.g. transcription of SME tacit knowledge) or semi-automated through any variety of processes (e.g. using unstructured texts or corpus of texts, natural language processing to identify nodes of interest, and artificial-intelligence classifiers). Creating the links and weights between nodes can also be done manually with small models or through semi-automated processes, such as those used to identify and populate the node sets. Examples of collecting data to support the existence of inter-nodal links include surveys of SMEs, sampling existing data from one or more sources, or processing Enron emails to establish who sent email to whom.

Once the first draft of the model is complete, the analyst will, preferably, have time to seek and gain objective and subjective validation of the model. When objective validation (e.g. ensuring nodes, links, and link values represent real-world data or fall within usually occurring ranges) is not feasible, subjective validation by SMEs is preferable to no validation at all. Ideally, the SMEs solicited to review the validity of the model would be different than the SMEs solicited for data to build the model in the first place.

With a validated model, the analysts are in a position to conduct static (i.e. single snap-shot in time) network analysis using traditional social network analysis for the agent-by-agent networks and node sets. As importantly, the dynamic network model supports analysis of the additional aspects of the world the model represents (e.g. who has access to what resources; what tasks precede other tasks; who is assigned what tasks and has or is missing the resources for those tasks; who was at what event(s). See also Appendix 3).

Insert the Model into an Agent-based Simulator

Also, with this model of nodes and links, the analyst is able to port the model to the Agent-based Model (ABM) simulator discussed in previous sections. Porting of the model can be completely automated or can go through some transformations that allow the analyst to stylize and generalize the specific model. An example of such transformations is converting IT-systems from resources to agents (assuming that was the original encoding scheme) to support the notion that humans can interact with and gain information from computers, databases, or other such systems. Such IT-systems-as-agents clearly have different cognitive, physical, and interaction parameters than the humans-as-agents. Another example would be the transformation of events into events-as-agents (aka special agents) that ensure knowledge associated with an event can be disseminated within the simulation. Like the IT-systems-as-agents, these special agents have very different parameters than humans-as-agents, while at the same time can be enabled/disabled to reflect the occurrence, re-occurrence, and post-event liveliness of the event's knowledge.

Analyze Outputs of Simulation(s) for Relevance to the Question of Interest and Hypothesis

Analysis of the simulations' output can support or negate the hypothesis, drive modifications to the hypothesis, and modifications to the simulations' parameters, as well as support sensitivity analysis to the modifications of the input parameters' groupings and interactions.

REQUIREMENTS

This method of modeling requires the analyst have and be familiar with at least two special-purpose capabilities, as well as office productivity software such as Microsoft's Office™. The two special-purpose capabilities are 1) software for computation dynamic and social network analysis (like CMU's ORA tool) and 2) simulation software (like CMU's Construct tool).

DATA

Average (3)

System dynamics data requirements are fairly small compared to other techniques. The analyst needs to know enough about a system structure, behavior and dynamics. Open-source information provides a wealth of cause-and-effect relationships and characterization of many aspects of historical behavior. Wargaming can provide insight into the thought processes underlying the domain and system, which affects how the model specification. Subject matter expert data gotten through formal and informal elicitation techniques can provide very valuable and tangible data regarding model structure, key behaviors and relationships, and system dynamics. Typically the data is quantitative in nature, and can be time-series, given many variable definitions in the system dynamic models are based on difference equations.

SET UP TIME

Short (2)

System dynamics can model many different time epochs, but they tend to best model longer term effects (such as tipping points) based on short-term stimuli. The time necessary to create a model is a function of the problem, the granularity and fidelity of the model, and the bounds of the model definition. Some models can be created in very little time (e.g., less than an hour), others can take weeks or months to define if not longer, it just depends on the level of complexity of the model.

EXECUTION TIME
Short (0.25)
Most models can be run (through simulation tools) in very little time (e.g., less than an hour).

SKILL SET/EXPERTISE
Specialized-Average (1)
System dynamics modeling does require a fair amount of training in the fundamentals and application of systems theory and modeling. There is an abundance of literature on the subject, as well as countless examples to help the modeler gain experience and understanding required to accurately undertake a system dynamics analysis. Some background in mathematics, programming and simulation is critical for setting up and solving more complex model structures. In addition, if there is a desire to compare the model to real world empirical data then it will also be necessary to incorporate domain expertise and experience in the design and conduct of experiments to test the model. The primary resource required to construct a system dynamics model is an experienced system theorist and modeler who can design a model that accurately captures the problem at hand and parameters of the system. SME's can be, and often are, a critical source of the system dynamics design.

TOOLS
Average (1)
Convenient system dynamics software and programs have been developed into user-friendly versions and have been applied to diverse systems. Examples include Stella, Vensim, AnyLogic, and countless other examples that can be found online.

COST
$ (7.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Going back to the 5D framework, discuss how the general questions identified in the 5D can be effectively addressed using this method.

- How effective is current US force posture for achieving policy objective?

This model allows the analyst to address a piece of this question; specifically, given the current US force posture, and therefore lines of alliance and "adversarialness" with other countries, how likely is it that the nations of interest will move in a direction in support or opposition of the US policy? In essence, that policy is expressed in terms of the pro/con attitude that the US desires the various countries of interest to evince, and then actions and information sharing by the US are the interventions evaluated to assess whether they will move the others closer or further from the desired state.

- What would be the optimal force posture to achieve a specific policy objective?

This model is not an optimization tool. It can be used to generate a suite of hypothetical futures that could then be assessed to identify those most in keeping with a specific policy objective. The outcomes of this model could be put into an optimization framework. To assess this question, a set of alternative force postures and the associated belief structures would be represented in the model. Then the interventions associated with those force postures would be run during a virtual experiment. The results would then be assessed to identify plausible futures of interest. It should be noted that enumerating all possible interventions of interest associated with a force posture is generally not feasible, so identifying an optimal force posture is less likely than is identifying the relative strength of alternative force postures.

- What strategy is optimal to achieve the objective?

To use these tools to answer this question, the analyst would need to specify a set of strategies, run a virtual experiment where these alternative strategies were run, then the results would need to be statistically analyzed. We note that, in general, optimality is often not the goal; rather, the objective is to identify a set of strategies that meet the objective so that factors external to the model can be used to choose between them.

Agent-based simulations are, by nature, abstractions of the real world. Agents in Construct have cognitive capabilities but are not humans. Construct agents have no emotional capabilities and have, presently, fairly limited goal/task-oriented capabilities. The simulations' results are therefore much more applicable to emergent and aggregated behavior analysis than specific agent behavior analysis.

Simulations that operate in multiple dimensions can be difficult to fully grasp for humans (technically-oriented humans, as well as lay-people). The very attempt to capture complexities of real life human interactions within the simulations can make communicating both the design and the output of the simulation difficult, potentially degrading the confidence of clients/customers in the results communicated to them by analysts.

ORA is being used to assess strategic nuclear capability of various countries. In addition, there is a project that just started to assess the relative influence of countries on each other choosing to develop a nuclear weapons capability using Construct.

Attempting to discern why a particular agent performed a particular way at a particular time is not a question Construct is equipped to answer with medium or high confidence.

Interactions of inputs, as well as parameters of the simulation, are almost always non-linear. Fine tuning a model, akin to changing an equalizer of a stereo system, can lead to discovery of inflection points and potentially discontinuities in outputs that are not immediately discernable, or anticipated, by analysts or clients/customers. As such, it is less likely that a question of "How much of X do I need to cause Y" will be easily answered in this model if X and Y link to other inputs or outputs.

FURTHER RESOURCES

The basic tools used are ORA for dynamic network analysis and Construct for agent-based dynamic-network computer simulation. The executable files, sample data, tech reports, and papers are available at these sites:

http://www.casos.cs.cmu.edu/projects/ora/
http://www.casos.cs.cmu.edu/projects/construct/

ORA has a built-in help feature. Additional sample data sets are available on the CASOS web.

Kathleen M. Carley, Michael K. Martin and Brian Hirshman, 2009, "The Etiology of Social Change," Topics in Cognitive Science, 1.4:621-650.

Kathleen M. Carley, Jeff Reminga, Jon Storrick, and Dave Columbus, 2011, "ORA User's Guide 2011," Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-11-107.

Brian R. Hirshman, Geoffrey P. Morgan, Jesse R. St. Charles and Kathleen M. Carley, 2010, "Construct Demo Input Deck," Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-10-118.

Jana Diesner and Kathleen M. Carley, "Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis," in Causal Mapping for Research in Information Technology, V. K. Narayanan and D. J. Armstrong, Eds., 2005, pp. 81-10.

Kathleen M. Carley, "Smart Agents and Organizations of the Future," in The Handbook of New Media, L. Lievrouw and S. Livingston, Eds., Thousand Oaks, CA, USA, 2002, pp. 206-220.

Lanham, Michael, Morgan, Geoffrey P., & Carley, Kathleen M. (2011, June). Data-driven diffusion modeling to examine deterrence. Paper presented at the IEEE Network Science Workshop, West Point, NY.

TIMED INFLUENCE NETS

Identify the levels, data observations and forms for which this method is most suitable.

Level

- ☑ **Global** (e.g., all global capable actors)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☐ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ❑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative and qualitative**

INTRODUCTION

Timed Influence Nets (TIN), a variant of Bayesian Nets, are used to capture cause/effect relationships that relate timed sequences of actions to the probability of an effect or outcome occurring. TIN models are thus well suited to capture the diverse aspects of nuclear strategy issues. Specifically, TIN models can be used to gain insights into the effects of actions on one or more nuclear strategy objectives and can be adapted to reflect different actors, international environments, phase of military operations, and scenarios. The TIN models can be enhanced through the use of multi-modeling techniques to leverage the ability of multi-agent-based modeling to capture the dynamic interactions among groups[31] and discrete event dynamical system models to provide insights into decision-making organizations.

METHODOLOGICAL OVERVIEW

Several modeling techniques are used to relate actions to effects. With respect to effects on physical systems, engineering or physics-based models have been developed that can predict the impact of various actions on systems and assess their vulnerabilities. When it comes to the cognitive belief and reasoning domain, engineering models are much less appropriate. The purpose of affecting the physical systems is to convince the leadership of an adversary to change its behavior, that is, to make decisions that it would not otherwise make. However, when an adversary is embedded within a culture and depends upon elements of that culture for support, the effects of physical actions may influence not only the adversary, but the individuals and organizations within the culture that can

[31] Details regarding Dynamic Network Analysis are provided separately; this paper only addresses Timed Influence Net models.

choose to support, be neutral, or oppose the adversary. Thus, the effects on the physical systems influence the beliefs and the decision making of the adversary and the cultural environment in which the adversary operates. Because of the subjective nature of belief and reasoning, probabilistic modeling techniques, such as Bayesian Nets and their Influence Net cousin, have been applied to these types of problems. Models created using these techniques can relate actions to effects through probabilistic cause and effect relationships. The adversary can use such probabilistic modeling techniques to analyze how the actions affect the beliefs and, thus, the support to and decisions.

Influence Nets (IN) and their Timed Influence Nets (TIN) extension are abstractions of Probabilistic Belief Nets, also called Bayesian Networks (BN) [2, 3], the popular tool among the Artificial Intelligence community for modeling uncertainty. BNs and TINs use a graph theoretic representation that shows the relationships between random variables. These random variables can represent various elements of a situation that can be described in a declarative statement, e.g., X happened, Y likes Z, etc. Pythia uses a Time Influence Net modeling extension to the Influence Net modeling paradigm.

An Influence Net (IN) is a Directed Acyclic Graph where nodes in a graph represent random variables, and the edges between nodes represent causal relationships. While mathematically IN are similar to Bayesian Nets (BN) [4], there are key differences. The most important is that IN uses CAST Logic [5,6] to enhance knowledge elicitation from subject matter experts in defining precise a-priori conditional probabilities used by BN. IN modeling is accomplished by creating a series of cause and effect relationships between desired (and undesired) effects and the set of actions that might impact their occurrence. The actionable events in IN are drawn as root nodes (nodes without incoming edges). Desired effects, or objectives the decision maker is interested in, are modeled as leaf nodes (nodes without outgoing edges). In some cases, internal nodes can be effects of interest, as well. Figure 1 shows an example IN. The actionable nodes (nodes on the left) have been assigned marginal probability values, a probability indicating whether or not an action will be taken. The other nodes have been assigned baseline probabilities using the CAST Logic, a probability indicating whether the random variable the node represents will be true on its own without the influence of any parent nodes in the model. The edges (links between nodes) have casual strength values (first two values) indicating the degree of influence that a parent node has on its child in the CAST Logic. The first strength value indicates the effect on the child node when the parent is 'True.' The second strength value indicates the effect on the child node when the parent is 'False.'

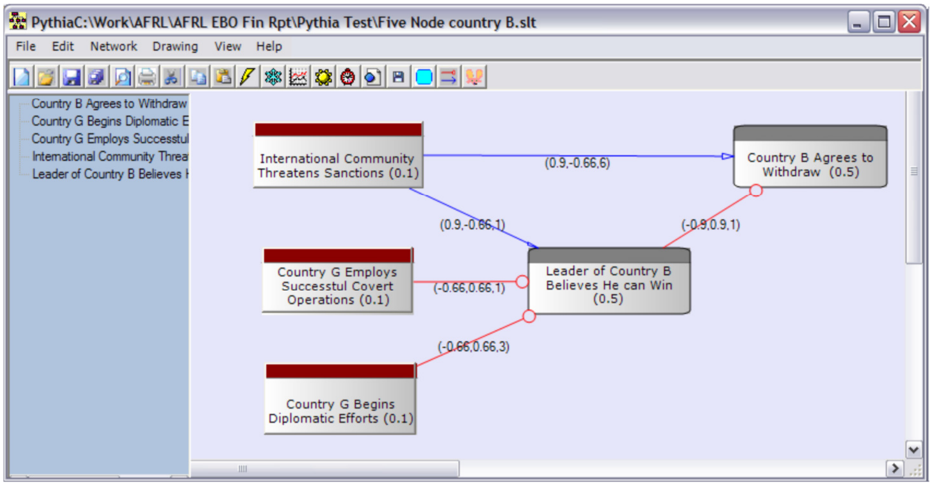


Fig. 1: Example Time Influence Net

The TIN modeling extension allows the modeler to allocate time delays associated with nodes and edges, representing an impact of events (actions or effects) that takes some time to reach and be processed by the affected events or conditions. For example, the third value assigned to edges (see Fig. 1) represents time delays in time units. Consequently, Time Stamps are associated with each node (including the action nodes). Hence, a user can specify a Course of Action (COA) as a time sequence on the action nodes, which are propagated through the network and trigger changes to the probability values of the effect nodes. The change in the probability value of a desired effect (leaf node) can be observed over time. Figure 2 illustrates the change in probability value for the effect node 'Country B Agrees to Withdraw' for the following COA on the action nodes:

> COA = ['Country G Begins Diplomatic Efforts' at time 1, and
> 'International Community Threatens Sanctions' at time 2, and
> 'Country G Employs Successful Covet Operations' at time 5]

Such a probability profile in TIN modeling provides important complementary information about the probability of success of a desirable effect when studied over time. First, the final probability level is given. Second, an unwanted drop in the probability level over time can be detected, and third, the time required to reach the final probability level is determined.

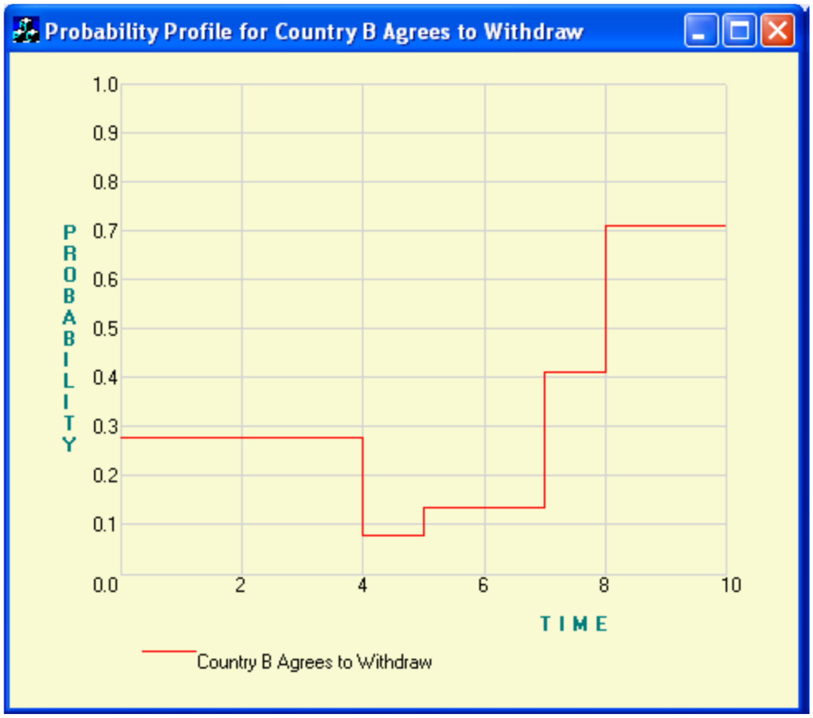


Fig. 2: Example Probability Profile

HOW IS THIS DONE?

Timed Influenced Networks graphically represent the analysis of cause-effect relationships for a question of interest. The model development process will highlight where gaps of knowledge exist; since the cause-effect relationships are stated in probabilistic terms, these gaps do not preclude exercising the TIN model to support the development and assessment of courses of action (COAs).

The basic steps of the process to develop a TIN model follow:

1. Clearly identify the question to be analyzed in terms of an overall desired objective.
2. Identify the major effects that contribute to the overall desired objective.
3. Determine key factors that contribute to achievement of the effect, as well as key factors that work against achievement of the effect.
4. Analyze the relationships between each of these key factors (positive or negative) to each of the major effects.
5. Determine supporting factors that contribute to (or detract from) achievement of the key factors.
6. Analyze the relationships between each of the supporting factors (positive or negative) to each of the key factors.
7. Continue to break down the supporting factors using the same cause-effect process until reaching basic causal elements that cannot be further decomposed.

8. Assign conditional probabilities for each link between cause and effect nodes. Two probabilities are required: The probability the effect node will be true if the cause node is true, and the probability that the effect node will not be true if the cause node is false. (These can be updated as new information becomes available to the analyst.)
9. Assign time required for an effect to be achieved from the time a causal node becomes "true."

To use the model:

1. Turn basic causal elements related to the given scenario "on" or "off" (these can be time sequenced).
2. Turn basic causal elements reflecting a desired course of action “on" or "off" (these also can be time sequenced).
3. Run the model and examine the probabilities associated with nodes of interest.
4. Conduct sensitivity analyses to gain insights into the relative contributions of individual basic causal nodes.
5. Conduct detailed time-probability profile analyses to refine COA development or gain detailed insights into the contributing factors related to the model outcomes.

REQUIREMENTS

To model using Timed Influence Net (TIN), models requires an understanding of cause-effect relationships relative to desired effects, knowledge of specific activities associated with a course of action, and an understanding of the relevant actor(s) core beliefs.

DATA

Average (3)

The data required for TIN model development varies based on the question(s) to be studied; however, for purposes of nuclear strategy analysis, the analyst will need an understanding of the key actors' core beliefs, behaviors, relationships, and force structure capabilities.

SET UP TIME

Average (3)

Defining the problem/issues to be studied, analyzing the cause-effect relationships, and collecting the necessary supporting data can be time consuming depending on the issue and the availability of supporting analyses to underpin the analysis of the particular problem under study.

EXECUTION TIME

Short (0.25)

Running the model and assessing the data can be done in a matter of hours, if the problem is properly defined. As an example, while it took several weeks to develop the initial generic model to support the CANS study, subsequent models in support of the wargames were designed and run in one or two days. With preparation, models can be used to examine excursions in a matter of hours during wargames or other such events.

SKILL SET/EXPERTISE
Specialized-average (2)
There is a rich behavioral influence community familiar with the use of Time Influence Net modeling and other forms of Bayesian Influence net modeling. They are widely used throughout the intelligence community to inform the analysis of many behavioral influence analysis problems. The use of multi-modeling does not find the same wide application, but is fairly intuitive for experienced members of the analytical community. The primary challenge is to build development teams, which collectively possess subject matter knowledge, modeling experience, and the necessary problem development and analytical skills to address complex issues. For analysis of specific regional issues, a second challenge is conducting research to provide data related to actor core beliefs and means to influence key stakeholders; this challenge is the focus of the Multi-agent-based Dynamic Analysis which is the subject of a separate paper. The models themselves are fairly easy to interpret, once developed.

TOOLS
Specialized (2)
Pythia (a time influence net modeling application/tool) requires only a standard x86 computer to operate. The program and associated files occupy approximately 33MB of storage and create an ".slt" file type that is based on XML standards. Pythia is readily available for download (http://sysarch.gmu.edu/main/software/), but because it is not certified for use on government networks, it requires use of a standalone computer in a DoD environment. No special licenses are required.

COST
$$ (10.25)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

Figure 3 illustrates the approach pursued to answer questions and provide analysts insight on issues related to both the "5D" framework and CANS "Vortex." The first step is to clearly identify the question the analyst intends to address as it relates to clearly-defined policy objectives. The Timed Influence Net and Dynamic Organization models can then be developed to examine the effect of policy or force structure changes on the ability to meet multiple objectives under varying environmental conditions and scenarios. For example, in phase zero, de-alerting ICBMs might seem a wise policy, but when the same cause-effect relationships are examined in a period of crisis, it suggests that the US would find it difficult to "re-alert" the ICBMs without escalating the crisis further. It is particularly useful as a means to understand the relationships of actions and effects such that actions taken to further one objective might have negative effects on other objectives. Typical questions include:

- What is the effect of actions taken to support one policy objective on other policy objectives of interest?

- What is the effect of establishing a force posture designed to support a policy objective in phase zero on the ability to support a specific policy objective in phase 1?
- What course of action (set of actions) is optimal to achieve multiple objectives?
- What is the mechanism that causes a seemingly positive action to have negative consequences?

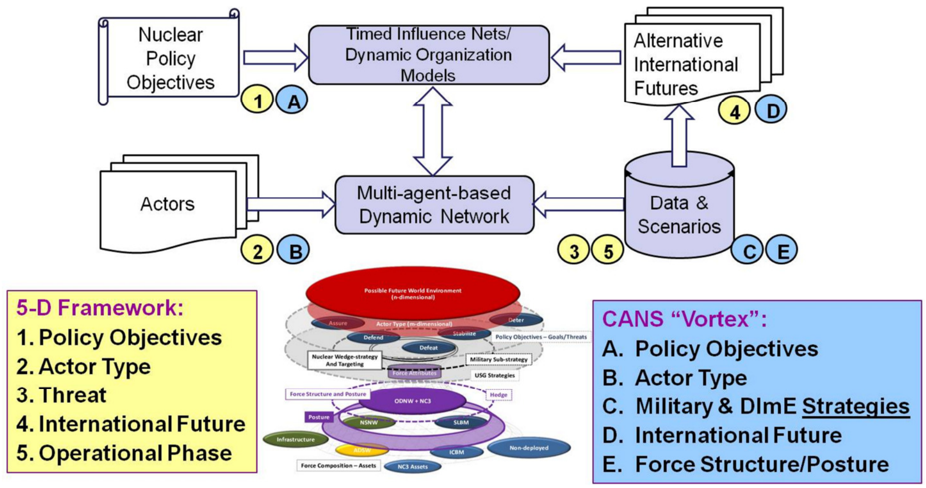


Fig. 3 TIN Model Relationships to CANS Frameworks

The use of multi-modeling provides a useful means to develop, analyze, and assess courses of action to address complex national security objectives, particularly in cases where actions taken to meet these objectives can cause positive effects relative to one objective and negative effects relative to another. It also provides a mechanism to track effects of actions in a variety of scenarios across multiple operational phases; similarly, these actions can be evaluated across multiple actors to highlight potential adverse consequences of actions and suggest mitigation opportunities.

FURTHER RESOURCES

References

Chang, K. C., Lehner, P. E., Levis, A. H., Zaidi, S. A. K., and Zhao, X., "On Causal Influence Logic", Technical Report, George Mason University, Center of Excellence for C3I, Dec. 1994.

Haider, S., and Levis. A.H., "Dynamic Influence Nets: An Extension of Timed Influence Nets for Modeling Dynamic Uncertain Situations". Proc. 10th International Command and Control Research and Technology Symposium, Washington DC, June 2005.

Jensen, F. V., Bayesian Networks and Decision Graphs, Springer-Verlag, 2001.

Kansal, S. K., AbuSharekh, A., and Levis, A. H., "Computationally Derived Models of Adversary Organizations," Proc. IEEE Symp. On Computational Intelligence for Security and Defense Applications 2007, Honolulu, HI, April 2007.

Rosen, J. A., and Smith, W. L., "Influence Net Modeling with Causal Strengths: An Evolutionary Approach", Proceedings of the Command and Control Research and Technology Symposium, Naval Post Graduate School, Monterey CA, Jun. 1996.

Zaidi, A. K., and Levis, A. H., "TEMPER: A Temporal Programmer for Time-Sensitive Control of Discrete Event Systems," IEEE Trans. on Systems, Man, and Cybernetics, vol. 31, No. 6, Nov. 2001, pp. 485-96.

WAR GAMING AND SIMULATION

Identify the levels, data observations and forms for which this method is most suitable. **Level**

- ☑ **Large n/global** (e.g., all internationally-designated terrorist groups)
- ☑ **Regional/ multi-actor grouping** (e.g., all South American countries)
- ☑ **Single nation-state or non-state actor**
- ☑ **Sub-national/organization group** (e.g., Pakistani military)
- ☑ **Individual decision maker** (Kim Jong-Il; President of Columbia, etc.)

Observations

- ❑ **Time-series:** multiple observations of the same actor or actors over time (e.g., monthly for the past 10 years)
- ☑ **Snap shot:** Fewer than three observations, or all observations occur at the same point in time

Form

- ❑ **Quantitative**
- ❑ **Qualitative**
- ☑ **Quantitative and qualitative**

INTRODUCTION

The DoD defines a war game as “a simulation, by whatever means, of a military operation involving two or more opposing forces, using rules, data, and procedures designed to depict an actual or assumed real-life situation” (*Department of Defense Dictionary*). Perla and Barrett (1985) define war gaming as “any type of warfare model or simulation, not involving actual military forces, in which the flow of events is affected by decisions made during the course of those events by ‘players’ representing the opposing sides” (p. 70). Both of these definitions indicate that war games serve to simulate realities of war in an attempt to gain a deeper understanding or knowledge of a particular scenario (Rubel, 2006). For centuries, simulations have been incorporated into military training with recent advances affording such training opportunities as learning piloting skills, tank maneuvering, team playing, and decision-making. Such exercises serve to provide a generic military understanding as well as to provide some specificity of potential upcoming operations (Rubel, 2006). The utility of war games and simulations is that they provide rich qualitative data from the exercise itself, enabling analysts and others to assess the mechanisms of decision-making.

METHODOLOGICAL OVERVIEW

War games and gaming have been around for centuries—the oldest and most well-known being chess. War games run along a continuum from basic strategy games (e.g., chess and “Go,” a less well-known game) to distillations “in which significant simplifications of reality are made for specific purposes" (Rubel, 2006, p. 114) to the most elaborate detailed simulations aiming to represent reality as much as possible. Chess provides the most basic illustrative example of what a war game can do while providing historical insight into why it itself is a war game. A strategic battle fought by two sides, chess is reminiscent of the kinds of battles fought in the early years of civilization: two armies literally facing each other across a field, organized by the general, with soldiers and archers in the front, and cavalry alongside. Over the years, chess and other war games evolved into more elaborate and detailed scenarios. For example, the Prussian Army in the early part of the 19th century began to develop and utilize more realistic war games "for training, planning and testing military operations" (Dunnigan, 2005, p. 141). In the twentieth century, prominent and telling examples of the use of war gaming comes from WWII Nazi Germany and the U.S. Gulf War in 1990. "When the allies invaded France on June 6th, 1944, the Germans were in the middle of a war game dealing with just such a possibility. As reality had overtaken the games hypothetical premise, the German commander ordered the game to proceed, but not as a game but as a command tool" (Dunnigan, 2005, p. 223; see also Snyder, 1989). Similarly, General H. Norman Schwarzkopf tells the story of the training exercise that eerily paralleled real life and provided insights and planning for what would become Desert Shield (see Schwarzkopf, 1992). It should be noted, however, that war games and simulations are not intended to serve as predictions or validations of future war plans as there are too many uncontrollable factors to be able to utilize war games and simulations in such a fashion. Nevertheless, the technique is valuable insofar as insights and understandings may be produced that would otherwise never see the light of day. If nothing else, war games and simulations may be able to reduce the element of surprise in the real world of military tactics.

War games are designed for both commercial (i.e., the average civilian on a home-entertainment system) and professional purposes (i.e., for military and governments to train or strategize,) or both. Commercial games will not be discussed further in this chapter. It is worthwhile to mention, however, that commercial games can be useful to understanding and designing war games, as some commercial products have been modified for professional (i.e., government) purposes. Moreover, with the rise of computer programming and availability of smaller and faster machines, better graphics and the like, computer games have become more prominent in the war gaming arena. It is important to note, however, that war games are not computer programs. What makes a war game is the human component—which may utilize computer programs to run a simulation or provide the virtual environment in which players can interact. Snyder (1989) notes that what makes a war game is "the psychology of combat—the determination and leadership of commanders at all levels—[that] really determines outcomes. Good war games emphasize these factors" (Snyder, 1989, p. 54). The benefit of playing a game or running a simulation is the larger picture it provides participants and analysts—often providing a new perspective on the sequence of events and decisions that are often missed while in the thick of action.

Rubel (2001) outlines five dimensions of military war gaming (see Figure 1 below). The first three levels form a solid foundation on which games are built, the last two provide higher level assessments that account for the modern advent of network-centric warfare (i.e., when information connects individual combat units):

1. Orchestration of one's own forces: The most basic and fundamental level of war gaming involves some orchestration of one's own forces that can be either physical markers on a board or map, or computer symbols that are moved about in some simulated fashion. It is possible to stop at this level.

2. Outcomes assessment: For training and strategy purposes, this second level is imperative. At this point players are confronted by other forces determined either by rolling dice, complex algorithms, or the game controller (or umpire). This is where inputs to the game or simulation (either by adversary players or by controllers) are assessed to determine their influence on subsequent moves in the game. This level provides the opportunities to gain important insights surrounding tactical play and decision-making.

3. Decision analysis: Level three is more critical to educational gaming situations than research games. With the third level comes decision analysis in which players are encouraged to "perceive objectively their own reactions to warfare situations" (Rubel, 2001, p. 64).

As games become more advanced, and life becomes more technologically oriented, network-centric warfare becomes more prominent. The final two levels account for such a style of warfare:

4. Shared awareness: "When units know what is going on and are confident that others do as well" (Rubel, 2001, p. 65).

5. Network behavior: This is how the different interconnected pieces on a network act via voice, video, and data transmissions, as well as any other type of communication necessary.

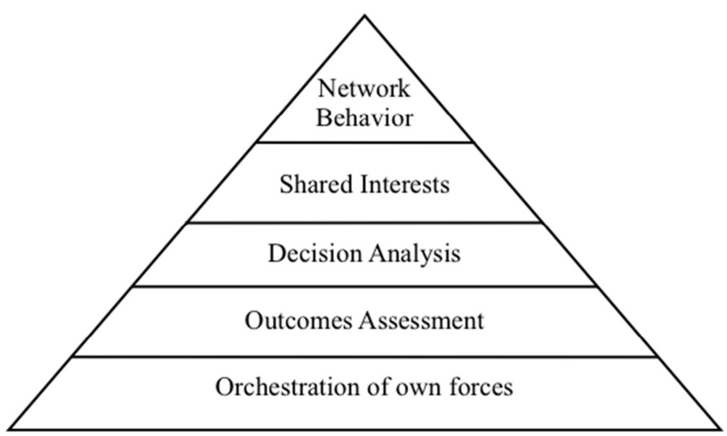


Figure 7. Basic and Higher-Level Dimensions of Gaming

HOW IS THIS DONE?

With the dimensions of gaming that are utilized, the main question that remains is: How would one go about designing a successful war game? War games require five sets of rules (Snyder, 1989, p. 50):

1. Objective rules (i.e., what is the end goal and how do players achieve it?)
2. Knowledge rules (i.e., what information is passed on to players and when?)
3. Execution rules (i.e., who controls the action? the players themselves or the "umpires"?)
4. Movement rules (i.e., how and when can a player move, and what factors limit movement?)
5. Engagement outcome rules (i.e., how is an outcome reached either among the controllers (umpires) or the players?)

Given this set of rules, designing a war game involves several key elements, described below, that must be addressed (see, Perla & Barrett, 1985; Snyder, 1989).

Purpose/Objective. First off, what is the purpose of the game? Will it be for educational/training purposes, or for research and strategy assessment? Along with knowing the purpose, one needs an objective. Objectives (or goals for the game) must be clearly defined and stated. Every other elemental design of the game must directly relate back to the game's purpose and objective.

Scenario/Set-up. The game scenario sets the stage for action. How many sides will play the game; one-sided versus the control, two-sided with teams playing against one another; or multi-sided in which many countries and agencies are included? Other scenario factors involve situational information including physical and environmental conditions, and other technical facts (e.g., the state of the weather, the forces available, a measurement of the forces' capabilities, etc.) (see Perla & Barrett, 1985). If the game is to be successful (or have a chance at success) much thought must go into the scenario so that it will realistically provide information to the players that will enable them to make decisions during play. Scenarios are a direct product of the game's purpose and objective. For example, a training/educational simulation with the objective of preparing pilots for flight in the Middle East desert must include the same weather patterns, geographical markers, and technical abilities in the scenario to be an effective (i.e., realistic) training simulation.

Models/Outcomes Assessment. Another aspect to the game is deciding how outcomes will be assessed. This can be either a top-down process where umpires (controllers) first make the calls and then go forward or a bottom-up process where small interactions are assessed by themselves then aggregated (see Snyder, 1989). Computers are more adept at bottom-up processing. Therefore, mathematical expressions can be built into the game that will respond to player moves and translate data and decisions into game events.

Rules/Procedures. Another critical element is for players to know what they can and cannot do, and why. For this element, it is important to determine whether the game will be "open" play (as in chess) or "closed" (as in life, or poker). In other words, can all players see all action at all times or does some player information remain secret until played? Additionally, players need to know where interactions are taking place; on a game board, with individual equipment, across networks, or among group platoons.

Levels of Play. Another important factor to consider is whether play will be at the tactical (or local), theatre (or operational), or strategic (or global) levels. Again, depending on the purpose of the game—education or research—the level of play will vary. In addition, the levels of play increase the time and intensity of the game—strategic level games require large amounts of people and time to implement, and are next to impossible to replicate. Other characteristics of each play level are outlined in Table 1.

Table 1. Levels of War Games (Perla & Barrett, 1985, p. 74)

	Local/Tactical	**Theatre / Operational**	**Global/Strategic**
Primary decision-maker	Battle group or lower	Commanders in chief	National command authorities
Goals	Give participants a better perspective. Compare various tactics and forces. Identify critical factors and areas for further study.	Explore specific issues. Identify strategic, operational and logistical problems in theatre. Identify areas for further study.	Give participants a better perspective, test a strategy; identify key issues.
Focus	Force levels and tactical deployments, weapon and sensor performance and interrelationships among warfare areas.	Necessary or feasible force levels and employment options for accomplishing specific military missions.	Pre-hostilities and transition politics and force deployments, the D-day shootout, escalation and war termination.
Primary output	Balance of qualitative and quantitative results. Number of iterations may vary, but tends to be higher than others.	Qualitative. Narratives and interpretations but with some numerical data for more support. Typically a small number of games run.	Qualitative. Narratives and interpretations with little numerical data; typically only a single game run.

Players. The final necessary element of any game is the players. This is crucial to war gaming because it provides that critical human aspect of the interaction between individuals and their environment. Players are typically assigned roles on various teams, such as the Blue Team (i.e., the home team, or the U.S.) or the Red Team (the adversary). In the early years of US Naval war gaming at the Naval War College, game plans were developed against specific countries that were represented by colors. The United States was always Blue, while, for example, Japan was Orange. Other games were known by the color-coded plans representing the other country: Red for Great Britain, Black for Germany; Silver for Italy. "The Maritime Strategy is the modern equivalent of the Orange plan, and the color Orange—no longer referring to Japan —is often the color of the major opponent, while Blue remains the color of the home team" (Snyder, 1989, p. 51). Today's color games have shifted from Orange adversaries to Red ones.

Additionally, modern games may involve other team players not directly involved in action and decision-making, but who are there to offer support and guidance, such as a team of subject matter experts (SMEs). This may be especially the case in network-centric gaming (NCW). "Whereas previously gamers would use tactical experts as umpires and analysts, in NCW gaming they may want to involve psychologists or other social scientists, as well as perhaps physiologists and physicians" (Rubel, 2001, p. 69). Finally, a Policy team may be incorporated into the game. In such a situation, players involved in the action make recommendations to the Policy team for consideration of current and future policies.

REQUIREMENTS

DATA

Low (2)

There are no a priori data available for use in war games or simulations in the traditional sense of data analysis techniques. Historical war case studies often serve as educational data inputs into modern war games. In addition, the selected topic around the desired objective must be researched, while any accompanying materials (such as audio, video, or print) must be developed. Despite the lack of traditional data analysis inputs, the game or simulation itself does provide a plethora of data that could be utilized after the simulation has played out, some quantitative, but most qualitative.

SET-UP TIME

Average (3)

War gaming is a fairly time intensive technique to set-up, in which a game designer must provide objectives, scenarios, equipment, etc.

EXECUTION TIME

Average (2)

The length of time to run the game depends entirely on the game itself: it can last for as little as a few hours up to as much as a few days, weeks, or even years.

SKILL SET / EXPERTISE

Specialized-average

Designing a successful war game or simulation takes experience and understanding of potential gaming artifacts that invalidate the game's findings. For instance, human error by game players, computer error, or even "invalid decision making by players" (Rubel, 2006, p. 115) due to player alienation, ignorance, or improper training, as well as the tendency for players to become overly aggressive due to the simulated reality can impact a positive conclusion to a game. "Game designers must therefore understand these tendencies and attempt to structure their games to minimize the likelihood and intensity of this player artifact" (Rubel, 2006, p. 116). The game designer must have (or acquire) knowledge of the issue or topic that will be the game's focus. Game controllers or umpires should also be well-versed in the topic; however, this need not be a requirement. Similarly, dependent upon the game as an educational tool or a research/strategizing tool, players may or may not have the background knowledge regarding the topic or expected moves to be made. The one factor that does require explicit knowledge and is somewhat debated in the literature involves casting the Red Team players. Some argue that Americans who play on the Red Team will still think "as Americans" thus diluting the opportunity for the Blue Team to play against an actual adversary (see, e.g., Snyder, 1989).

TOOLS

Minimal (0)

War gaming and simulations require a number of tools dependent upon the game being played. At its most basic, this technique requires people who are either subject matter experts (SMEs) or military personnel (e.g., soldiers, pilots, lieutenants, commanders, etc.). The more elaborate the game or simulation, the more tools are needed, including computers, computer programs and virtual environments, and physical space.

COST

$ (8)

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

War gaming and simulations are useful in "developing questions, issues, and provisional insights" (Watman, 2003, p. 53) by producing indicative knowledge. In other words, games and simulations afford an opportunity for players to visualize relationships and connections that might not normally be seen. Taking this a step further, if the game objective is for strategy or research purposes, the game as it was played can be dissected for more information. Analyzing or assessing the game can be done in three steps (see Rubel, 2006).

1. Research the facts
2. Trace the effects back to their causes
3. Evaluate the means employed

These three steps "establish a criterion for the extraction of valid knowledge from a war game" (Rubel, 2006, p. 117).

Two types of game assessment are possible: rigid and free. As Rubel (2006) describes, rigidly assessed games "proceed strictly according to rules governing movement, detection, and combat … the game goes where the rules take it; if the rules and the combat-resolution tables are good representations of reality, the outcome constitutes artificial military history, and one can usefully work backward from outcomes and look for reasons" (p. 118). Freely assessed games, on the other hand, are run by umpires and game directors. Rather than following the rulebook, the game's progress is "governed by the objectives of the game's sponsors, the time available, and sometimes the conflicting interests of stakeholders" (Rubel, 2006, p. 118).

Notably, "there are certain warfare problems that only gaming will illuminate" (Rubel, 2006, p. 111). Although it is tempting to think that war games and simulations offer predictive indicators of future events, it is important to understand that this is simply not possible. Simulations and games are meant to *mimic* reality, but they can never *be* reality. "Fundamentally, war gaming is an experiment in human interaction and is best used to investigate processes, not to calculate outcomes" (Perla, 1987, p. 44). Thus, there will always likely be some factors that will remain unforeseen. To drive this point home, consider that a war game can never be run with exactly the same outcomes, regardless of whether the players remain the same. This is because humans experience shifts in mood, behavior, reactions, and other elements that will impact their choices during the game. Games and simulations thus provide special insight, including the possibility of uncovering unintended consequences. But such "findings" cannot be used as a proxy to predict outcomes in future real-world situations or events. Instead, they offer analysis and assessments of the operational environment, as well as the opportunity to visualize the full spectrum of adversary capabilities and intent. At their best, games and simulations provide an understanding of potential threat courses of action and a basis for intelligence direction and synchronization.

FURTHER RESOURCES

Bracken, P. & Shubik, M. (2001). War gaming in the information age. *Naval War College Review, Vol. LIV,* 47-60.

Dunnigan, J. F. (2005). *The Complete Wargames Handbook.* Retrieved from: http://loakes.game-host.org/limeyyankgames/games/RESOUR~1/COMPLE~1.PDF

Department of Defense dictionary of military and associated terms. (no date). Retrieved from http://www.dtic.mil/doctrine/dod_dictionary/.

Macedonia, M. (2002, March). Games soldiers play. *IEEE Spectrum,* 32-37.

Oswalt, I. (1993). Current applications, trends, and organizations in U.S. Military simulation and gaming. *Simulation & Gaming, 24,* 153-189.

Perla, P. P. (1987). War games, analyses, and exercises. *Naval War College Review, 40,* 44-52.

Perla, P. P. & Barrett, R. T. (1985). What wargaming is and is not. *Naval War College Review, Vol. XXXVIII,* 70-78.

Rubel, R. C. (2001). War-gaming network-centric warfare. *Naval War College Review, Vol. LIV,* 61-74.

Rubel, R. C. (2006). The epistemology of war gaming. *Naval War College Review, Vol. 59,* 108-128.
Schwarzkopf, H. N. (1992). *It doesn't take a hero.* Bantam Publishers.
Snyder, F. (1989). What is a war game? *Naval War College Review, Vol. XLII,* 47-54.
Watman, K. (2003). War gaming and its role in examining the future. *The Brown Journal of World Affairs, Vol. X,* 51-61.

EXAMPLES of APPLICATION of MODELING TECHNIQUES to CANS PROBLEM SPACE

DNI WARGAMING

POC: Dan Flynn, DNI, daniejf@dni.gov

HOW THIS SPECIFIC EFFORT FITS WITHIN THE GENERAL TECHNIQUE

DNI's gaming effort fits into the general technique by providing insight into a broad range of issues related to nuclear strategy. Rather than focusing on a highly specific element of US nuclear policy, it brought together a collection of experts from across the IC, DOD, academia and policy communities in order to examine how changes in policy will affect other nations. Gaming tests specific hypothesis to educate policy makers on potential "what ifs" of a given scenario. This DNI effort worked to place individuals into a strategic role and asked them to think more broadly about the impact of US nuclear policy changes. The conclusions drawn from discussions and actions among these experts provided insights into the discrete reactions of major friendly and un-friendly actors, giving a robust and comprehensive picture of the likely outcomes of potential changes in policy and posture.

SPECIFIC STRENGTHS OF THIS TECHNIQUE FOR THE STUDY OF NUCLEAR STRATEGY

Changes in nuclear posture and policy do not occur within a vacuum. US nuclear policy must be considered within the context of its repercussions across the international stage in a dynamic environment. The DNI strategic gaming technique provides insights into the wide ranging, and often unexpected, implications of how changing posture and policy will affect bilateral and multilateral relationships. It also provides a longer timeframe by examining not only the initial responses of actors, but analysis of extended changes to the status quo through several iterations of responses and counter measures adopted by pertinent actors.

Strategic gaming allows analysts to apply a wide range of scenarios for possible future world environments, allowing decision makers to determine effective responses to short-term and long-term changes in nuclear status quo. Utilized in the context of the Nuclear Posture Vortex[32], this flexibility can be used to examine future changes in policy objectives, nuclear sub-strategy, military sub-strategy, force structure and posture, and to some extent, force composition.

HOW YOUR WORK INCORPORATES AND/OR EXAMINES ELEMENTS OF THE 5D FRAMEWORK

Policy objectives- During each move players were asked to answer tasker questions that provide insights into the overall game objectives; these included specific elements of deterrence, assurance,

[32] For explanation of the NPV, see the CANS Theory Team Report.

defense, defeat, and strategic stability. At the conclusion of the game, participants were asked to participate in a hotwash, during which they evaluated lessons learned in the previous moves through an American mindset. Because teams must consider all of the variables in determining effective responses to changing game dynamics, the game necessarily provided comprehensive insights into constructing successful US policy in the strategic nuclear environment.

Actor type – Analysis through gaming allows for development of the qualitative analysis of various actors on the international stage. As stated in the draft CANS framework report, important factors for this analysis include military capability, international integration and territoriality, all of which are considered when conducting a game. Furthermore, the game is capable of providing insight on all sixteen of the sub-variables identified in the report under the broader headings of identity, relative power, relationships, WMD capability, and vulnerability. The DNI game included several teams of experts that each took on the role of different actors, allowing for detailed analysis of each actor along these variables.

Threat – The versatile nature of the game allows for close examination of threats associated with nuclear weapons and global security. The game was structured to allow analysts to determine pertinent issues to national security and the effects of calculated responses to those threats on the international stage as the United States reduces its nuclear forces. Games often allow for players to be surprised by the threats perceived by other teams in the game as policy actions are played out.

International future – Gaming is most important to the international future dimension. Gaming provides not only immediate implications, but a number of iterations and moves from actors; it gives useful insight for the effects that nuclear strategy will have on the world stage and vice versa. In the DNI futures framework, the game gave the capability to draw conclusions for a future scenario of return of great power confrontations, concert of powers, fragmented international system, rise of non-state networks, and most importantly, a future that contains elements of all four futures.

Phase – The DNI game gives analysts the ability to examine all stages of the operational phasing model and can provide insight on effective strategies to shape policy for Phase 0.

HOW YOUR WORK CONTRIBUTES TO OUR UNDERSTANDING OF THE ELEMENTS OF NUCLEAR STRATEGY.

In contrast to the service games, the DNI game is strategic in scope. The purpose and concept for the game was to gain insight into global responses to changes in US force posture, focusing on the iterative and complex nature of this relationship. The principal objective of the game was to understand that dynamic as the United States proceeds towards new and evolving nuclear posture.

The understanding gained from the DNI game will also allow for very specific examination of changed dynamics. The insights that are gained can be focused into narrowly-defined categories that are important to the US policies regarding nuclear weapons. For example, the game is capable of highlighting issues of deterrence, stability, etc. in the context of a singular bilateral relationship with either allies or opponents and then places those findings within a larger context of a potential world future. These findings can then be used to determine appropriate force attributes, posture, and other elements of nuclear strategy.

THE EXTENT TO WHICH YOUR FINDINGS CAN BE GENERALIZED

The findings of this DNI gaming effort can be generalized to include a wide range of issues related to nuclear policy. These include game-generated insights regarding the strategic objectives of actors, the way nuclear issues are perceived amongst the actors, and actor's perceptions concerning arms reductions.

USAF REGIONAL WARGAME METHODOLOGY

POC: Darphaus Mitchell, L CIV USAF AF/A10-A Darphaus.Mitchell@pentagon.af.mil

The Assistant Chief of Staff, Strategic Deterrence and Nuclear Integration, Headquarters, US Air Force (AF/A10) hosted the first in a series of planned wargames 19 -21 July 2011. The wargame supported independent, but complementary, Air Force and CANS objectives.
The Air Force game focused on producing operating concepts and courses of action (COAs) responsive to challenges presented by nuclear adversaries and on developing Air Force intellectual capital to begin the process of crafting a comprehensive 21st century deterrence, assurance, and stability vision that builds on the Nuclear Posture Review.

WHAT TYPES OF QUESTIONS CAN BE ANSWERED?

HOW THIS SPECIFIC EFFORT FITS WITHIN THE GENERAL TECHNIQUE

Using a seminar-styled tabletop wargaming method, the USAF wargame explored concepts for dealing with the complex issues of deterrence, assurance, and stability in a multi-polar, highly proliferated 21st century environment. The game was played by three Blue Teams comprised of a mix of functional and operational expertise from the Air Force, US Navy, US Army, USPACOM, and USSTRATCOM. In reaction to a challenging strategic environment, game players reacted to a postulated "road-to-war" series of events played out over three moves designed to test multiple phases of a regional campaign. Blue players were placed at the operational level of war as members of the Joint Force commander's (JFC) and appropriate component command strategy and planning staffs.

SPECIFIC STRENGTHS OF THIS TECHNIQUE FOR THE STUDY OF NUCLEAR STRATEGY

Wargaming method allows players the opportunity to productively role play within a highly interactive environment with time constraints and an interacting opponent. Players, through these interactions, were able to derive a richer experience and produce higher fidelity concepts of operation and courses of action (COAs).

HOW YOUR WORK INCORPORATES AND/OR EXAMINES ELEMENTS OF THE 5D FRAMEWORK

Policy objectives. The game assumed a continuation of current policy objectives through the 2020 timeframe. Policy objectives played were deterrence, assurance, strategic stability, and defeat. Given the scope of our tabletop, we did not actively game the counter-proliferation objective.

Actor types. The game explored the motivations, behavior and actions of adversary, allied, and neutral actors through game play. While Red, Green and White teams' interactions were scripted, their actions were unpredictable, and the game scenario fully examined the military component of the DIME construct.

Phases. Blue players were placed at the operational level of war as members of the Joint Force commander's and appropriate component command strategy and planning staffs. Game scenario moves were designed to gain insight into player perceptions and actions across a range of campaign phases from shaping to decisive operations. Blue team players independently developed COAs appropriate to each move and briefed the JFC in plenary session each day. Briefings included rich discussions on Blue perceptions, assumptions, and conceptualizations of allied tasks; operational risk; limiting factors; forces required; unintended consequences; mitigation strategies; concepts of operation; employment plans; key assumptions regarding adversary objectives, allied objectives, neutral party objectives; team-developed evaluation criteria (e.g., proportionality, risk, effect on deterrence/assurance/stability, risk of escalation, timeliness). COAs were evaluated in terms of how they fit into national policy objectives and overall strategy.

HOW YOUR WORK CONTRIBUTES TO OUR UNDERSTANDING OF THE ELEMENTS OF NUCLEAR STRATEGY

The game sought to provide game players the opportunity to address the complexities presented by nuclear weapons in the context of a future regional conflict; their role in deterring conflict; the strengths, weaknesses and synergies of conventional, nuclear, cyber, ballistic missile defense, and space capabilities; and the operational implications of alternative force postures, alliance and adversary capabilities.

EXTENT TO WHICH YOUR FINDINGS CAN BE GENERALIZED

While USAF regional game findings can be generalized over multiple components of the 5-D framework, the game was designed to address operational level issues and the connections to strategic level policy objectives. Like the overall CANS approach, the game was not structured to answer nuclear force structure or posture questions, or to recommend specific force capability solutions. However, concepts of operations developed included various COAs tailored to the game's actors, objectives, capabilities, and actions. Additionally, the game produced a number of insights for deterrence strategists and campaign planners that could have broader applicability (with appropriate tailoring in accordance with unique regional dynamics) to other geographic combatant commanders.

MONITOR 360 SME CROWDSOURCING FOR SOCIO-CULTURAL INSIGHTS

POC: Derek Bothereau, Monitor 360, Derek_Bothereau@monitor-360.com

For the CANS effort, Monitor 360 conducted SME crowdsourcing as an analytic technique to gather socio-cultural insights related to nuclear strategy. These socio-cultural insights were synthesized and delivered to CANS in the form of Key Segments, Master Narratives, and Implications for Wargame Planners. Monitor 360 conducted SME interviews, research, and analysis on Japan, South Korea, and China for this effort, incorporating both country-specific and regional perspectives from over 30 SMEs.

KEY SEGMENTS

Each country section begins with a summary of the "Key Segments" of decision-makers and influencers in that country, providing readers with insight into different clusters of opinion on key regional security issues. The "Key Segments" are identified through SME interviews and analysis.

- The concept of "Key Segments" is a recognition that the views of decision-makers and influencers are **not monolithic** in any country and are, in fact, much more complex and varied. Key Segments (such as the "Pro-U.S. Establishment" in Japan) are comprised of a number of **major players** in the political and military establishment, as well as members of the public and opposition groups.
- In some cases, members of an entity, such as a political party or a government agency, may exhibit characteristics of more than one Key Segment, which speaks to the inherent complexities and contradictions that exist. After identifying the major players in each segment, the full report describes the **core beliefs of that segment across a range of critical issues**, including regional security, relationships with the U.S., and the role of nuclear weapons.

MASTER NARRATIVES AND SUB-NARRATIVES

The report also identifies Master Narratives and Sub-Narratives for each country, which helps explain *why* views among Key Segments exist and how they are derived from culturally and historically rooted experiences. Master Narratives and Sub-Narratives depict the relatively "timeless" cultural and historical lenses through which actors in the region will interpret events and policies today and in the future. Master Narratives are a useful tool for understanding how segments of non-U.S. populations understand who they are and where they come from, and how they make sense of developments unfolding around them. Such insights are critical for understanding the decision-making environment in which actors operate.

- **Master Narratives** operate at a macro level and are the historically grounded stories that reflect a community's cultural identity and experience, or explain its hopes, aspirations, and concerns.

- **Sub-Narratives** stem from Master Narratives and focus on more specific regional issues. Understanding these multiple Sub-Narratives can provide U.S. government planners and decision-makers with critical context for the views that drive concerns, aspirations, and policy agendas in East Asia.

IMPLICATIONS FOR WARGAME PLANNERS

Finally, the report draws implications for wargame planners from the analysis of Key Segments and Master Narratives and Sub-Narratives.

- Each Key Segment in Japan, South Korea, or China is likely to view future "moves" by regional actors, as well as by the United States in a different light.

- Understanding these different segments, and the lenses (narratives) through which they will interpret regional developments, can provide added insight for wargame planners who are considering how future actions and scenarios could reverberate throughout the region.

Using findings from SME crowdsourcing, as well as knowledge of likely high-impact events in the region, **Monitor 360 developed sample moves** (defined as adjustments in security policy or diplomatic relationships in the region) to demonstrate how this material might be used in a wargame.

Extent to which these insights can be generalized: *This wargame section in the report represents* ***several hypothetical illustrations****. Wargame planners will have other moves and specific scenarios that they may want to test in the region, as the specific conditions that define the future environment may facilitate different responses from foreign actors than those portrayed in the report.*

THE ATTRIBUTE TRADE-OFF MODEL (ATOM)

POC: Tom Allen, NSI, Inc. tallen@nsiteam.com

ATOM relies on an assessment process that first analyzes a problem structure from complex concepts to more basic and directly measurable elements and then synthesizes the evaluation of those basic elements through the structure so that alternatives may be assessed not only on the basics, but on the high-order concepts as well. The first challenge raised by this task is determining how to link discrete and measurable force posture attributes (such as flexibility, sustainability and reach) to such broad concepts as deterrence and counter proliferation in a systematic and meaningful way. ATOM achieves this by creating a theoretical model that decomposes these high-level policy objectives into their basic elements, and then links individual force posture attributes to these specific elements. The theoretical model draws on an extensive academic and policy literature to determine the set of elements for specific policy objectives.

The second challenge is to derive assessments with respect to high-level concepts such as policy objectives from the evaluation of the more basic elements of the model decomposition such as force posture attributes. There are many algorithms designed to aid in this process—what is often referred to as multi-attribute decision analysis—and ATOM includes two that have been instantiated into its software. A fuller description of these algorithms appears below in the ATOM Software Overview section of this document.

The ATOM software is composed of two parts: (1) a Java-based Structure Authoring Tool that provides users a graphical interface for decomposing the problem space and; (2) An R-based Decision Support Engine (DSE) that aggregates the assessment of force posture alternatives through to policy objectives, cost and risk. In essence the software takes the model and represents it graphically in the form of tree diagrams that clearly map the breakdown of individual policy objectives and the link between policy elements and force posture attributes. This relational information is then used by the DSE to assess the relative strengths of specific force postures for achieving individual or multiple policy objectives.
ATOM, as presented in this guide, therefore, should be thought of as two related, but distinct products. The first is the theoretical model, which is specific to the nuclear policy context; the second is the software, which, although developed to deal with this specific model, is in itself content-free. The Structure Authoring Tool and DSE can be used to render a detailed decomposition and analysis of any problem space of interest to the analyst, from nuclear policy to which motorcycle to buy. It is our expectation that for analysts interested in the nuclear policy problem space there will be very little need to change the current instantiation of the theoretical model. Two possible exceptions to this would be modifications of the edge weightings (which are currently all set at 1.0, implying equal weighting of each child node) and additional linkages between specific policy elements and force posture attributes. The majority of input will be done in the DSE, with the comparison of specific force postures (represented by their ratings across the 13 meta attributes taken from STRATCOM's existing analysis structure) across different combinations of policy objectives.

ATOM NUCLEAR POLICY SPACE MODEL

The ATOM nuclear policy space model starts by identifying the top level components of US nuclear policy most commonly referred to in policy and doctrine: deterrence, assurance, defeat, counter proliferation and strategic stability. For the purposes of this model, however, further refinement of these objectives was necessary. First, strategic stability and counter proliferation are considered as higher-level goals, achieved through the application of a specific policy: deterrence, assurance or defeat. Furthermore, policies of deterrence, assurance and defeat are considered in the policy and academic literature to be context dependent and thus, must be further defined. Deterrence or assurance designed to counter proliferation have different requirement dimensions from those of deterrence or assurance for strategic stability. It is also clear from the academic literature that deterring an attack against one's own territory is a different problem from deterring attack against a third party. Finally, policy statements and concept papers make a clear distinction between defeat designed to neutralize an opponent's military capability and defeat with the intent to destroy. To these refined policies were added cost and risk, which figure prominently in the DO JOC and are intrinsic to STRATCOM's current analysis process.

Nine top level nuclear policy components:

1. Direct deterrence
2. Extended deterrence for strategic stability
3. Extended deterrence for counter-proliferation
4. Assurance for strategic stability
5. Assurance for counter-proliferation
6. Defeat (neutralize)
7. Defeat (destroy)
8. Cost
9. Risk

The ATOM model starts by mapping the disaggregation of individual policy objectives, then linking, where possible, individual force posture meta-attributes to those elements. The alternative approach would have been to start with the list of force posture meta-attributes and determine how these may influence specific aspects of a policy objective. This is significant, as it means that the ATOM model includes policy elements that are not (as far as we know) directly affected by force posture attributes. Thus, the model tells the analyst not only where force posture can make a difference to achieving a policy objective, but also where it cannot. This can be of particular interest in situations where tradeoffs between policy objectives arise. Figure 1 illustrates the complete ATOM nuclear policy space model. The inner circle (blue) comprises the nine top-level components, the next ring (red) are the policy dimensions, the third ring (green) the elements of those dimensions and the outer ring (white) the force posture attributes associated with specific policy elements.

The rationale behind the disaggregation and specification of each of these policy components will be discussed next. This discussion is not an exhaustive review of the literature behind the ATOM model; rather, it is intended to walk the user through each branch of the ATOM hierarchy. A brief definition of each policy objective is given for each branch, then an explanation of how each of its component dimensions and elements are defined. The primary purpose of this explanation is to ensure that the analyst has a clear understanding of the scope of explanation that the ATOM nuclear policy space model incorporates. Once the theoretical component of the model is explained, the logic behind the connection of the force posture meta-attributes (see **Error! Reference source not found.**) to the theoretical portion of the model is discussed. This will then enable the analyst to better interpret the evaluations of force postures generated by the software and place their analysis within a strong theoretically driven policy problem space.

The ATOM model is designed to examine how well a specific force posture can contribute to a set of policy objectives, relative to other possible force postures. While the theoretical model is fully specified at the conceptual level, it is not fully operationalized. The only measured attributes that are incorporated into the DSE assessments are those relating to force posture. This is a critical distinction for the analyst to keep in mind when interpreting the results generated by the DSE. These assessments indicate how well a particular force posture can contribute to a specified set of policy objectives. It is not an overall assessment of how well the US will be able to achieve its policy objectives. Since the ATOM theoretical model includes attributes that are external to force postures, future implementations of the ATOM model may include estimates of values on these attributes to study how exogenous factors may bear on the way different force postures are ultimately assessed.

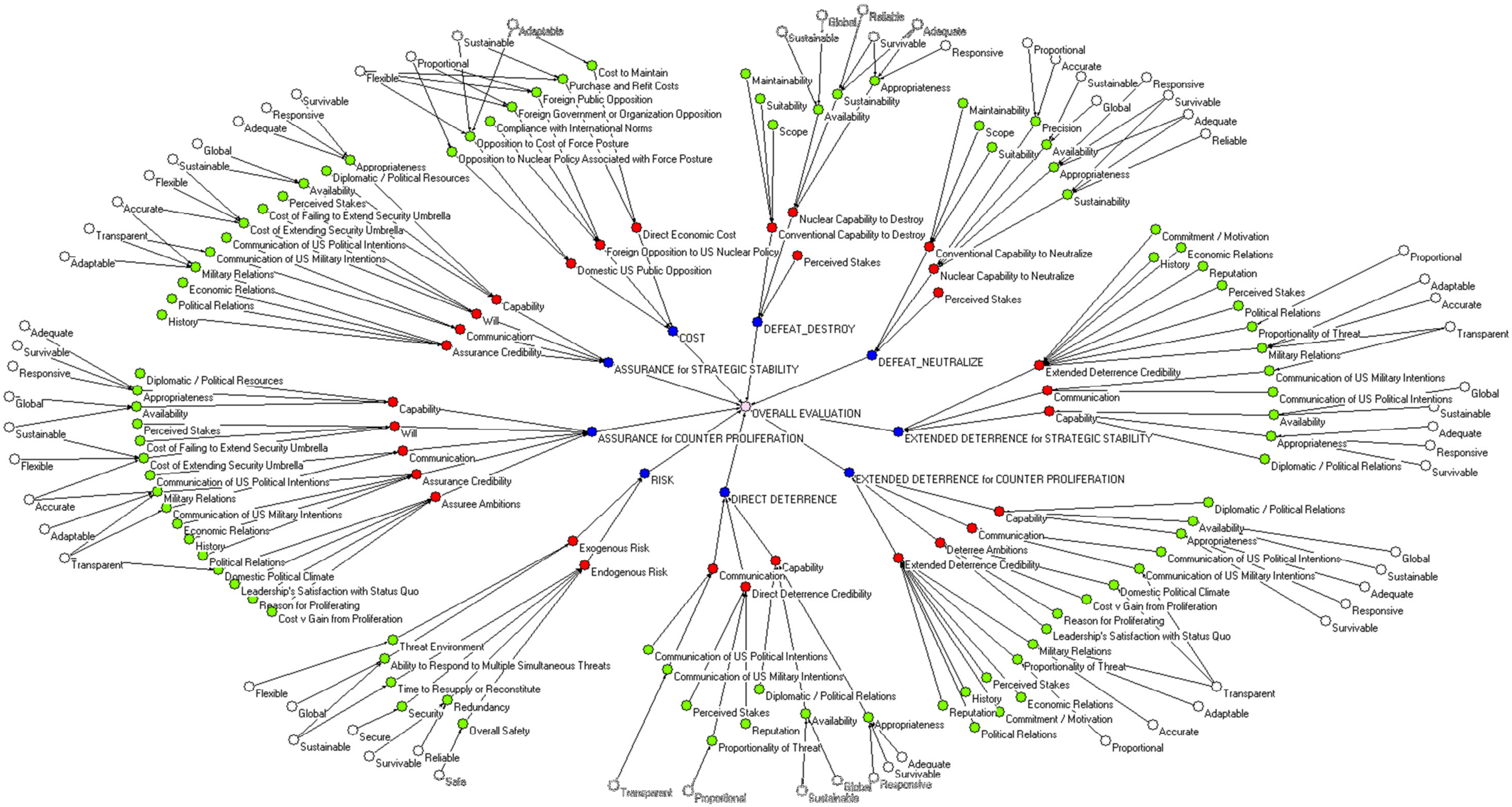


Figure 1, Entire ATOM nuclear policy space

ATOM SOFTWARE OVERVIEW

The ATOM software is comprised of a Structure Authoring Tool to assist users in decomposing complex problems into a hierarchical structure, terminating in basic attributes that can be directly assessed, and a decision support engine that aggregates the evaluation of higher-order concepts in the decomposition on the basis of their contributing elements.

ATOM STRUCTURE AUTHORING TOOL

The ATOM Structure Authoring Tool allows users to configure and visualize the relationships between policy objectives and force posture meta-attributes that comprise the problem space. The Structure Authoring Tool exports a structure file for use in the ATOM Decision Support Engine (DSE).

Functionality

- ATOM visualization can be used independently as a tool for mapping a particular problem space, or the structure can be exported into the DSE for further analysis.
- Users can manipulate and change: elements (policy objectives and dimensions, force posture meta-attributes); relationships between elements; and the direction and weights of these relationships.
- Important unmeasured elements can be included in visualization yet excluded from computational analysis (gray nodes in Figure 8).

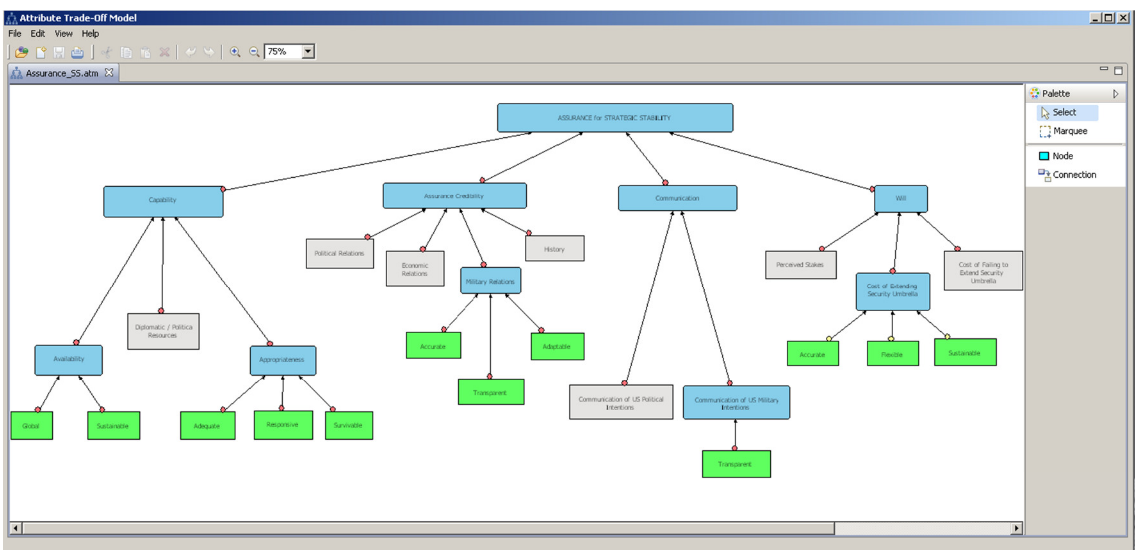


Figure 8. Screen shot of the ATOM Structure Authoring Tool

ATOM DECISION SUPPORT ENGINE

The DSE reads in a problem structure and assessment values for the force posture meta-attributes, which collectively constitute a **force posture** for the purposes of the ATOM analysis. It then aggregates the assessments up the structure using either the Simple Additive Weighting (SAW, see Yoon & Hwang 1995: 32-36) or Evidential Reasoning (ER, see Yang & Xu 2002) method. The DSE allows users to compare as many alternative force postures as desired. It also permits users to specify as many alternative combinations of high-level policy objectives (the top-level branches of the problem space hierarchy) as desired to obtain an overall assessment. For example, users may wish to examine an overall assessment considering only Assurance for Strategic Stability and Cost, or they may wish an overall assessment considering all forms of Deterrence and Risk.

Aggregation algorithms

SAW: Aggregates by assessing a parent as the weighted average assessment of its descendants.

- Weights are the edge weights supplied in the structural decomposition of the problem space and specified in the ATOM Structure Authoring Tool.
- Assessments at the most basic level of the decomposition are on force posture meta-attributes and each meta-attribute must be given a single value within the discrete evaluation scale.
- Parent nodes may not be evaluated to a value on the discrete evaluation scale, but may be assessed to intermediate values (see Figure 9).

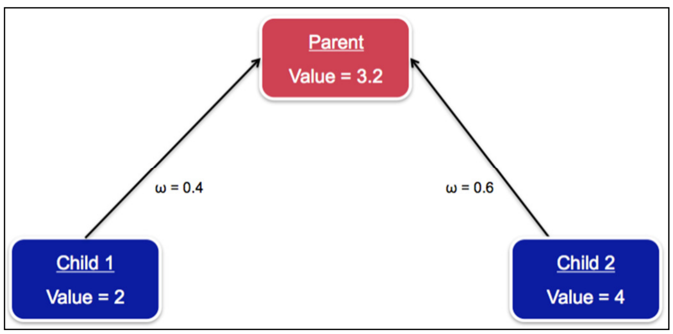


Figure 9. Example aggregation using SAW

ER: Employs a more complex scheme whereby assessments of elements may be distributed across the evaluation scale rather than concentrated on one value.

- For example, rather than assessing an element as a 2 out of 5, one could assess it as a 2 with 75% *belief* and a 3 with 25% belief (see Figure 10).
- Parent nodes in the decomposition are assessed on the same scale as their descendants, with the distribution of belief determined by belief distributions on the child nodes and the edge weightings joining them to the parent.

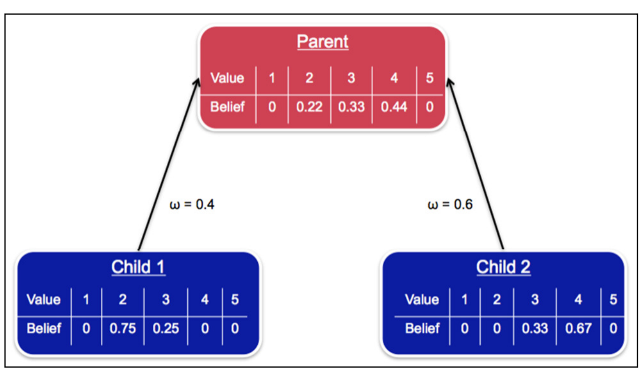


Figure 10. Example aggregation using ER

Comparison of aggregation methodologies

Simple Additive Weighting

- Simple to use and explain
- Requires evaluating the most basic elements to an exact value on the assessment scale
- Assumes that the assessment scale is interval-valued; that is, the true value between any two points on the assessment scale is assumed to be the same
- Calculated nodes will usually not be expressed on the same scale as the basic nodes; that is, they will be fractional whereas the basic scale may be integer-valued
- Can lend a false sense of precision to the decision analysis process

Evidential Reasoning

- Not as simple to use or explain
- Allows for uncertainty in the assessment of basic elements, including the type of varied input that might be obtained by surveying a panel of experts
- Results of the assessment aggregation may require additional assumptions regarding the utility of each value on the evaluation scale to conclude a definitive ranking of choice alternatives
- Well suited to identifying alternatives with extreme risk/reward potential

APPLICATION OF THE TIN MODEL TO CANS

POC: Alex Levis, George Mason University, alevis@gmu.edu

Figure 1 outlines relationships between the TIN model and scenarios used to address the nuclear issues posed in the CANS study. The models were developed to address the five dimensions in the CANS "5-D" model: Policy objectives, Actors, Future environments, Threat, and Operational phase. The Time Influence Net model (created using the software implementation "Pythia") is used to decompose policy objectives into high-level effects; these effects are subsequently analyzed to identify actions and sub-actions which would contribute or detract from achieving these high-level effects. In general, the effects reflect perception of a decisionmaker or decisionmaking organization. One set of actions is derived from the scenarios to reflect the threat and phase; these generally run counter to the desired effects. Other actions reflect a course of action (COA) proposed to influence the actor to behave favorably relative to the desired effects. Subject matter experts work with Time Influence Net modelers to assign conditional probabilities to the cause-effect relationships, as well as estimate the relative start time for inputs (actions taken to influence the actor, create effects, and achieve desired objectives), and the time delay for each action to contribute to an effect or sub-effect. WebTAS, a visualization tool developed by AFRL and used extensively throughout the joint intelligence community, is used to display key events as a function of location and time.

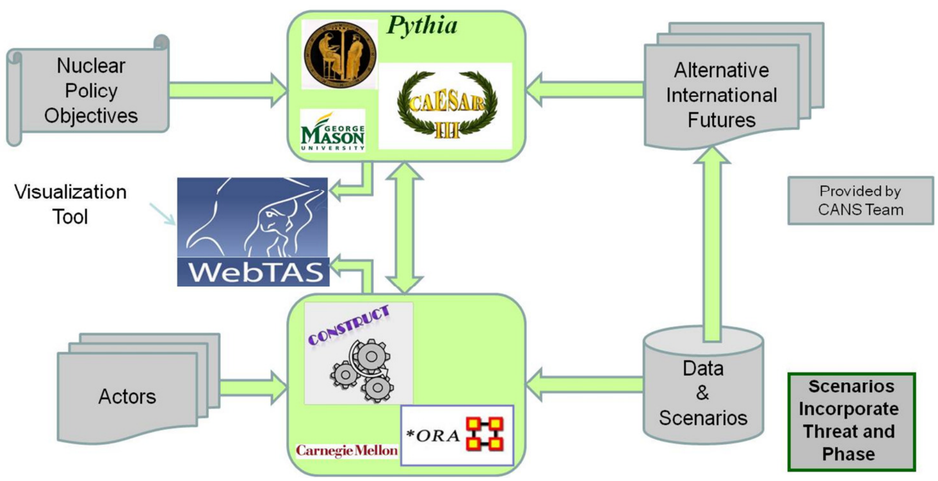


Fig. 1: TIN Model (Pythia) relationships for nuclear issues posed by CANS

Based on the work of the CANS theory team, a basic model was developed (See Fig. 2) starting with four key nuclear policy objectives: (a) Prevent proliferation of nuclear weapon technology and capabilities; (b) Assure friends and allies; (c) Deter potential adversaries; and (d) Maintain global and regional strategic stability. The model did not address the question of defeating a nuclear competitor except in the context of deterring that competitor from pursuing nuclear-related behaviors that might threaten US national interests. This led to development of a generic model based on examination of cause-effect relationships involving both potential adversaries and friends during the development of a crisis, with a primary focus on nuclear-related actions derived from traditional military strategy. Since most non-military strategy elements vary significantly based on actors and regions, these elements were described only as diplomatic, informational, or economic. Core beliefs are treated as scenario inputs for the generic Timed Influence Net model (Pythia); these can be refined when specific actors are provided during the examination of regional issues using regional template models.

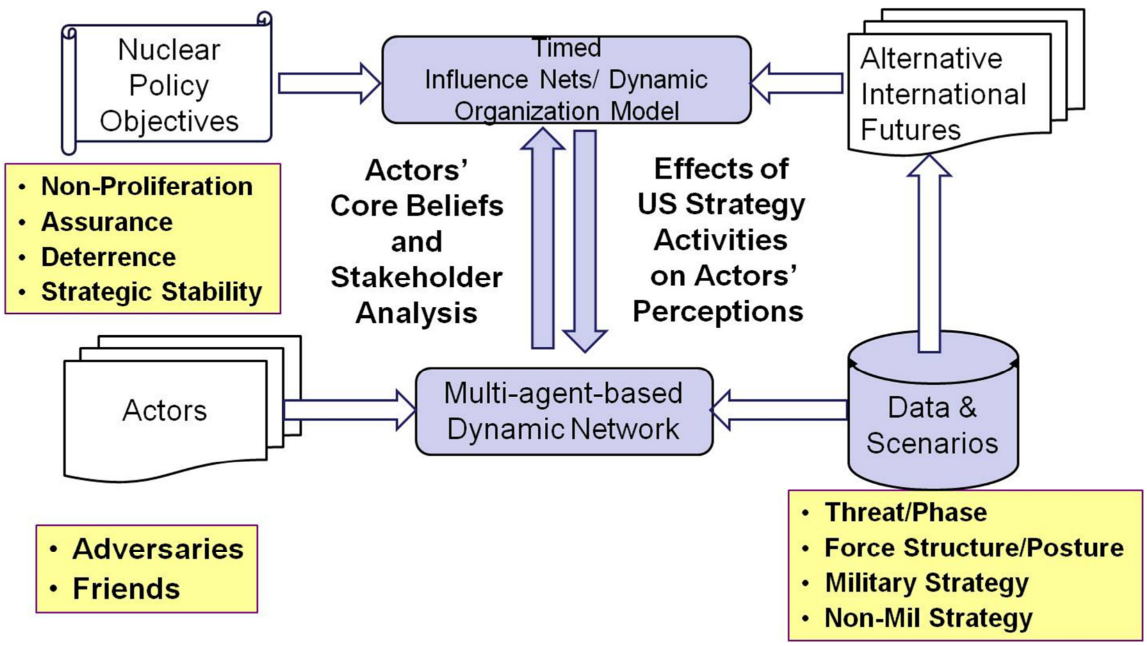


Fig. 2 CANS Multi-modeling Process

The generic model is a useful way to gain an understanding of basic cause-effect relationships, but the analysis of specific issues requires the development of issue-focused models. Such analysis begins with a determination of the objectives based on the question(s) to be answered, the scenario, and the actors. The two wargames conducted in support of the CANS project require development of specific regional models to meet the analytical needs surrounding the issues of interest; however, each subsequent model can build on previously-developed models with the top-down "generic" and bottom-up "specific regional issue" models converging at the intermediate " template regional " model (see Fig. 3).

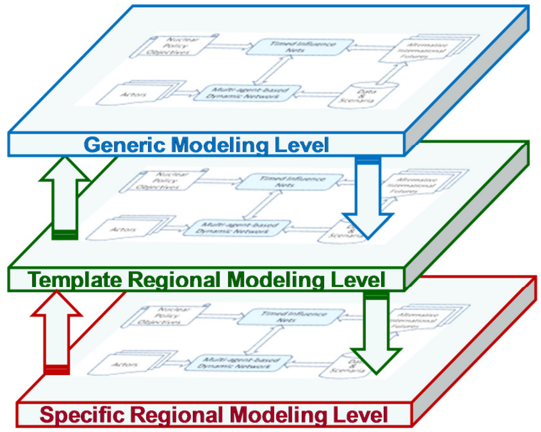


Fig. 3 Modeling Taxonomy

The models used in the CANS study reflect the basic cause-effect relationship outlined in the Deterrence Operations Joint Operating Concept (DO-JOC), which is depicted graphically in Fig. 4.

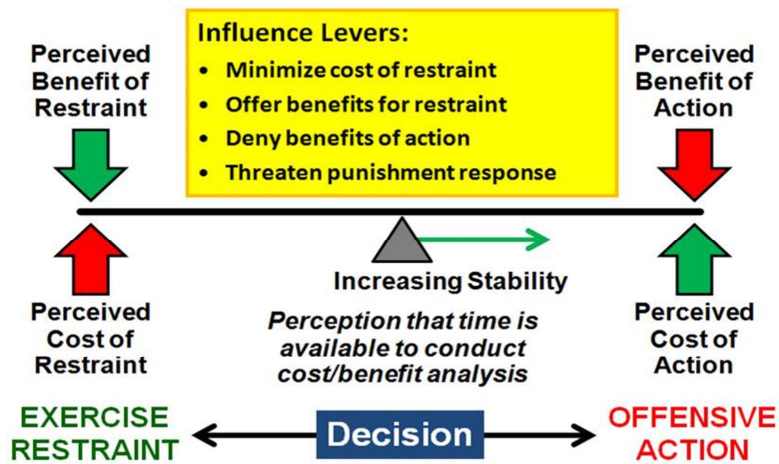


Fig. 4 Deterrence Operations Joint Operating Concept (DO-JOC)

Most of the data used in the TIN models supporting the CANS effort are qualitative in nature (See Fig. 5); however, they are characterized as causal strength parameters, which are then converted to conditional probabilities using the heuristic conversion algorithm which is fundamental to the Pythia implementation of Timed Influence Net model; so, although qualitative, the variables can be processed using quantitative methods.

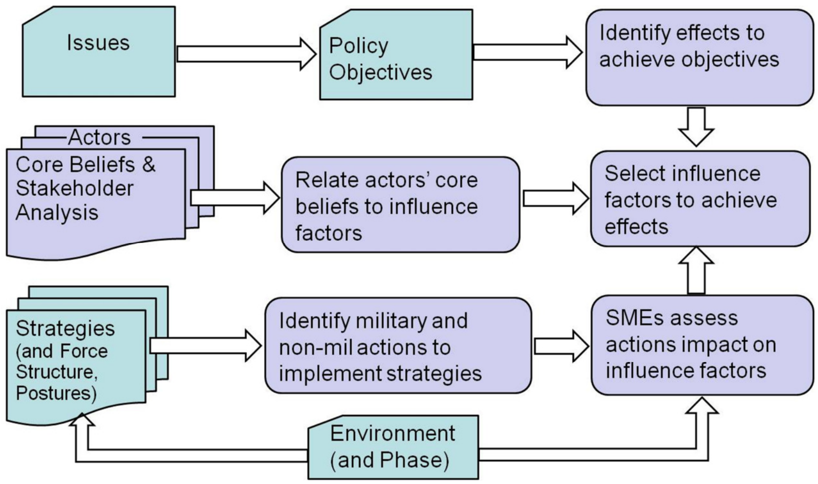


Fig. 5 CANS TIN Modeling Data Requirements

The input variables reflect all known actions that will influence the effects, whether they are based on the scenario, an actor's core beliefs, or an activity conducted as part of a course of action to achieve a desired effect. These variables are time sensitive; therefore, it is important to know the time stamp (relative time of occurrence) of scenario and response activities, as well as the time required for causal activities to create their effect. This analysis requires subject matter expert assessments with other models (such as the multi-agent based dynamic model) used to provide specific actor/stakeholder inputs.

The "Generic" TIN Model developed for CANS and later refined for use in two nuclear games (one sponsored by the Air Force and the other by the Navy) is depicted in Figure 6. The model consists of four major elements: (1) Objectives, (2) Effects of actions on perceptions related to the objectives, (3) Actions to influence perceptions in element two, and (4) actor core beliefs. Specific results of the two games cannot be provided here due to the classified nature of the game; however, in both cases, the generic model was adapted to a template regional model (two different regions were studied), and then further refined to specific regional models based on scenario requirements. With the benefit of the experience gained from analyses conducted following the Air Force game (which was held first), the model was used to provide insights to the participants of the Navy game during the course of the event.

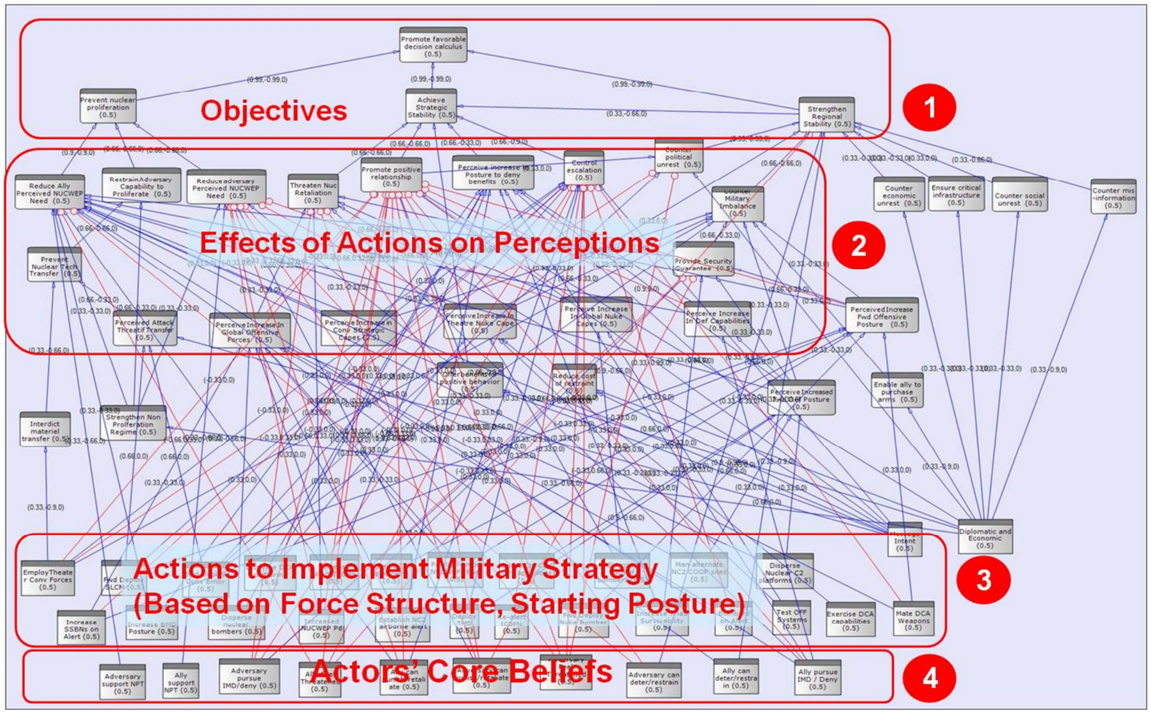


Figure 6 CANS "Generic" TIN Model

THE DECISION ANALYSIS TOOL (DAT)

POC: Tom Allen, NSI, Inc. tallen@nsiteam.com

The Decision Analysis Tool (DAT) is a structured framework for analyzing adversary decision making without assuming a single, static decision process or choice rule. It is designed to help the analyst quickly construct a decision matrix and assess the robustness of possible outcomes and assumptions about a decision actor's choice processes by providing results calculated according to a variety of choice rules. The matrix formalizes the calculus of a decision actor by identifying interests and perceived outcomes that reflect the decision actor's perception of the problem space, not an objective mapping of the problem space, or the analyst's own perception of the problem space.

The DAT decision matrix represents the decision space from the point of view of a specific decision actor. In the context of CANS the decision actor would be a nation, sub-national decision authority (e.g., executive or military elite), non-state actor or individual decision maker whose motivations US analysts would like to explore and understand. The decision actor modeled in DAT could be either an adversary (potential or actual) or an ally. The other actor modeled—what we have termed the response actor—is most often the US, however, this could also be some third party interacting with the decision actor in a scenario US analysts would like to examine.

Decision theory has shown that there is an implicit and unavoidable tension between the desire to achieve a good outcome and cognitive effort required to reach a decision. How decision makers solve this trade-off between effort and accuracy is one of the primary determinants of the decision model they employ. Quality, quantity and order of information, the nature of the response and the complexity and importance of the problem are all contextual factors that can affect the adoption of a decision model. Thus, not only can we not assume a single decision rule across all strategic contexts, we cannot assume that an individual decision maker will always employ the same decision model.

Decision models are important because they help us understand why individuals make the choices they do, and they can help us understand the effects of context on the outcome of strategic interactions. More important in the CANS context, perhaps, the choice of decision model can influence the outcome of the decision process—the choice of action. Thus, assuming a priori that all decision makers, in all situations, will employ the same decision model limits our analysis and potentially skews our results.

THE DAT DECISION MATRIX AND ALGORITHMS

DAT provides the following capabilities: Creation and editing of a decision matrix; and initialization of parameters and computation of a preferred outcome based upon several decision choice rules.
Creation and editing of the decision matrix
The users enter the following information:

- A set of decision options for two actors: the decision actor (D) and the response actor (R). We'll refer to these sets as ΩD and ΩR, respectively.
    - ΩD: actions available to the decision actor.
    - ΩR: actions the decision actor believes to be available to the response actor.

Jointly, these two sets form the set of outcomes, Ω, formally defined as: Ω = ΩD × ΩR. We can consider the set of outcomes as action/response pairs that are evaluated against a set of objectives for the decision actor, termed decision dimensions. We will denote the number of outcomes by n.

- A set of decision dimensions, Δ, that are the objectives or goals relevant to the decision actor. We will denote the number of decision dimensions by d.

Users enter this information into an n × d matrix on an Excel worksheet and enter preference values for each decision dimension in the cells of the matrix. For each decision dimension, the preference values rank each outcome against the others with respect to how well that outcome satisfies the dimension. Users may enter any values they wish, and, for each dimension, the system will standardize them to positive values in the range [1, n] that sum to and preserve the original rank order.

The layout of the DAT is shown in Figure . The decision matrix is on the left hand side, and all algorithm results are to the right. Rows within the primary layout correspond to inputs or results for a given outcome. Below the primary layout are inputs that relate to the items directly above, e.g., weights for the Decision Dimensions, fitness threshold for the Satisficing algorithm.

DPRK-US Alternative		Decision Dimension (Higher is better)						Statistics		Outcome Parameters	
DPRK	US	Relations w/ PRC	Relations w/ ROK	Economic	Regime Survival	National Security	Intl Prestige	Mean	Standard Deviation	Likelihood	Expediency
Test	Diplomatic	4	1	7.5	7	7.5	6.5	5.58	2.37	0.10	7
Test	Sanction	2	3.5	2.5	3.5	7.5	6.5	4.25	2.04	0.20	7
Test	Strike	3	3.5	5.5	5.5	4	6.5	4.67	1.25	0.30	5
Test	Invade	1	2	2.5	1.5	2	6.5	2.58	1.81	0.40	4
No Test	Diplomatic	5	5	7.5	8	6	2.5	5.67	1.82	0.40	3
No Test	Sanction	6	6	2.5	3.5	5	2.5	4.25	1.49	0.30	7
No Test	Strike	7	7	5.5	5.5	3	2.5	5.08	1.77	0.20	2
No Test	Invade	8	8	2.5	1.5	1	2.5	3.92	2.94	0.10	1
	Weight	0.10	0.14	0.09	0.18	0.22	0.27				

DPRK	US	Simple Normative SCORE	Simple Normative RANK	Weighted Normative SCORE	Weighted Normative RANK	E{Utility} SCORE	E{Utility} RANK	Weighted E{Utility} SCORE	Weighted E{Utility} RANK	Satisficing SCORE	Satisficing RANK	Lexicographic Choice SCORE	Lexicographic Choice RANK	Elimination by Aspects SCORE	Elimination by Aspects RANK
Test	Diplomatic	33.50	2	35.29	1	3.35	7	3.53	7			33.3%	2		
Test	Sanction	25.50	5	29.70	3	5.10	6	5.94	5						
Test	Strike	28.00	4	29.46	4	8.40	2	8.84	2	28.00	1			100.0%	1
Test	Invade	15.50	8	18.42	8	6.20	4	7.37	3						
No Test	Diplomatic	34.00	1	31.86	2	13.60	1	12.75	1			66.7%	1		
No Test	Sanction	25.50	5	24.42	6	7.65	3	7.32	4						
No Test	Strike	30.50	3	27.00	5	6.10	5	5.40	6						
No Test	Invade	23.50	7	19.85	7	2.35	8	1.98	8						

Satisficing		Lexicographic Choice	Elimination by Aspects	
Fitness Threshold	0.33	Ordering	Fitness Threshold	0.33
Cutoff Rank	2.64		Cutoff Rank	2.64

Figure 1. DAT Layout (DPRK Example)

ALGORITHMS

There are currently seven DAT decision rules: four Normative variants, Satisficing, Lexicographic Choice, and Elimination by Aspects.

Normative

Normative rules account for all objectives by using a sum across objectives for each alternative, with alternatives ranked in accordance with this sum.

Simple Normative

Uses a straight sum across objectives to score each alternative.
Inputs: Decision matrix

Weighted Normative

Takes a weighted sum across objectives (decision dimensions) to score each alternative, where objectives may not be uniformly weighted. The weightings are assumed to reflect priorities of the decision actor.
Decision dimension weights are entered directly below the decision matrix. Entered values must be in the range [0, 1]. The algorithm will normalize the inputs so that the weights sum to 1.
Inputs: Decision matrix, decision dimension weights

Expected Utility

Alternatives are initially scored as with the Simple Normative, but then each of these scores is multiplied by the respective likelihood of that alternative occurring. The likelihoods are assumed to reflect the judgment of the decision actor.

Likelihoods are entered in the Likelihood column (under Outcome Parameters). Entered values must be in the range [0, 1].

Inputs: Decision matrix, outcome likelihoods

Weighted Expected Utility

Alternatives are initially scored as with the Weighted Normative, but then each of these scores is multiplied by the respective likelihood of that alternative occurring. Weightings and likelihoods are again assumed to reflect the priorities and judgment of the decision actor.

Decision dimension weights are entered directly below the decision matrix. Entered values must be in the range [0, 1]. The algorithm will normalize the inputs so that the weights sum to 1. Likelihoods are entered in the Likelihood column (under Outcome Parameters). Entered values must be in the range [0, 1].

Inputs: Decision matrix, decision dimension weights, outcome likelihoods

Satisficing

A decision maker using a satisficing rule will select the first outcome satisfying some threshold condition.

Alternatives are ordered in terms of expediency, that is, an assessment of how the readily decision and response actors would each be able to implement their action options, respectively. Higher values of expediency indicate greater ease of implementation.

A threshold value is set and the first alternative, in terms of expediency, that has matrix entries all exceeding the threshold will be the preferred choice. A threshold is a percentile that each of an alternative's scores must achieve in order to be selected. For example, a threshold of 0.33 requires that a satisfactory alternative has none of its decision dimension rankings falling in the lower third.

The cutoff rank displayed below the fitness threshold is a calculated value and is provided for information purposes. Any alternative with an decision dimension rank below this cutoff value cannot be selected.

Inputs: Decision matrix, outcome expediency, fitness threshold

Lexicographic Choice

Identifies the outcomes associated with all possible (relevant) dimension prioritizations. A preference ordering is established on the set of decision dimensions, assumed to reflect the judgment of the decision actor. The alternative ranking highest on the first decision dimension in the preference ordering is chosen. If two or more alternatives are tied on this dimension, the alternative ranking highest on the next dimension in the preference ordering is selected. The process continues in this fashion through the preference ordering until all ties are broken or all the dimensions in the preference ordering are exhausted. If more than one alternative remains, then all of these are deemed selected.

DAT iterates the lexicographic choice algorithm over all permissible permutations of the decision dimensions, where permissibility is defined by order constraints that the user can impose. For example, a certain dimension may always need to be ranked first or one dimension may always need to be ranked ahead of another. DAT ranks alternatives on the basis of how often they are selected using the lexicographic choice algorithm over the permissible permutations of dimensions.

Elimination by Aspects (EBA)

EBA eliminates the outcomes associated with the worst outcome pay-off. As with lexicographic choice, first a preference ordering is established on the set of decision dimensions reflecting the judgment of the decision actor. In addition, a threshold value (percentile) is set. Alternatives not meeting the threshold on the first dimension in the preference ordering are eliminated. Of the remaining alternatives, those not meeting the threshold on the next dimension in the preference ordering are eliminated. The process continues in this fashion through the preference ordering until one alternative remains or all dimensions in the preference ordering are exhausted. If more than one alternative remains, then all of these are deemed selected.

DAT iterates the EBA algorithm over all permissible permutations of the decision dimensions. Permissibility is again defined by order constraints that the user can impose. DAT ranks alternatives on the basis of how often they are selected using the EBA algorithm over the permissible permutations of dimensions.

The cutoff rank displayed below the fitness threshold is a calculated value and is provided for information purposes.

Elimination by Aspects uses the same set of partial order relations as Lexicographic Choice.

Note: The run time of EBA may be considerable if the number of decision dimensions d is increased. Run time for the DPRK example (d = 6) is on the order of a few seconds. Increasing d to 8 will increase this by approximately a factor of 56 (= 7 × 8).

Inputs: Decision matrix, partial ordering, fitness threshold

DYNAMIC NETWORK ANALYSIS AND MODELING FOR CANS

POC: Kathleen Carley, Carnegie Mellon University, kathleen.carley@cs.cmu.edu

GOAL

Provide an analytical basis supported by computational modeling and simulation for addressing complex nuclear strategy issues.

BASIC APPROACH

CMU developed generic and region-specific models for use in multi-agent dynamic-network information and belief diffusion simulation framework, used Dynamic Network Analysis techniques and tools to analyze results of computer simulations, and accepted and provided feedback loops into GMU's Pythia Model.

SPECIFIC STRENGTHS

The approach we use has several specific strengths for the study of nuclear strategy. First, the approach is reusable. The generic model can be readily and rapidly adapted to a wide range of scenarios by just changing the set of actors. Second, subject matter information can be easily incorporated by setting the component actors within each country or non-state actor, the links among them and their tendencies to adhere to their national narratives, and follow human-interaction patterns and motivations. Third, this technique lets you identify emergent leaders, assess the relative rate at which various actors will move toward or away from a nuclear threat situation, and it can be used to identify the points of influence for defusing the entire situation.

USING THE CAN'S "VORTEX" AND 5D FRAMEWORK

Modelers at CMU built a generic multi-agent model using the five (5) key elements from the CANS vortex: Policy Objectives; Actor/Agent Type; Military and diplomatic, information, military, and economic (DIME) strategies; International Future and Force Structure/Posture. The generic model also incorporates the 5D framework established by the Theory Team: Policy Objectives; Actor Type; Threat; International Future and Operational Phase.

Modelers, at the behest of the Department of the Navy (DoN), developed representations of countries that would facilitate viewing adversaries, allies, as well as non-aligned entities as non-monolithic entities. Having multiple agent types representing each country allowed us to simulate various factors that affect countries' views of and reactions to the world they occupy. This representation technique also supported the ability to aggregate a non-monolithic entity (e.g. a country, a multi-national organization) into a single agent.

GENERIC MODEL DEVELOPMENT

CMU, with Subject Matter Expert (SME) advice and assistance, developed the following eight (8) generic agents for each country that could be included in the model: Executive; Legislative; Intelligence Community; Economy & Trade; Internal Affairs; External Affairs; Military; and Public Opinion. Limiting the number of key agent types to eight allowed depiction of conflicting motivations and policy objectives that most countries in the world must contend with, while reducing the challenges associated with attempting to gain information and intelligence to build higher fidelity (e.g. more actor types) models.

CMU established the generic model from the position of assuming that a decision-maker was interested in possible impacts to second and third parties, rather than direct impacts on the US itself. Therefore, in the generic model, the US is not modeled, except through it actions upon other actors. Because of the need to keep the model generic, we posited an ally and adversary actor of indeterminate nationality. We modeled the interconnections of the Ally actor through the layering of multiple Erdös-Renyi graphs and used a Core-Periphery network to define the adversary actor. The networks were weighted, and those weights had implications for the flow of information between actors – links had 1 of 4 possible weights: no influence (the absence of a link), little influence, moderate influence, and high influence. We assume that most future models would include a representation of the United States, which makes the generic model US-centric while remaining flexible to represent any regional/geographic area of interest to the US government (USG). An ORA representation of this network is shown in Figure 2.

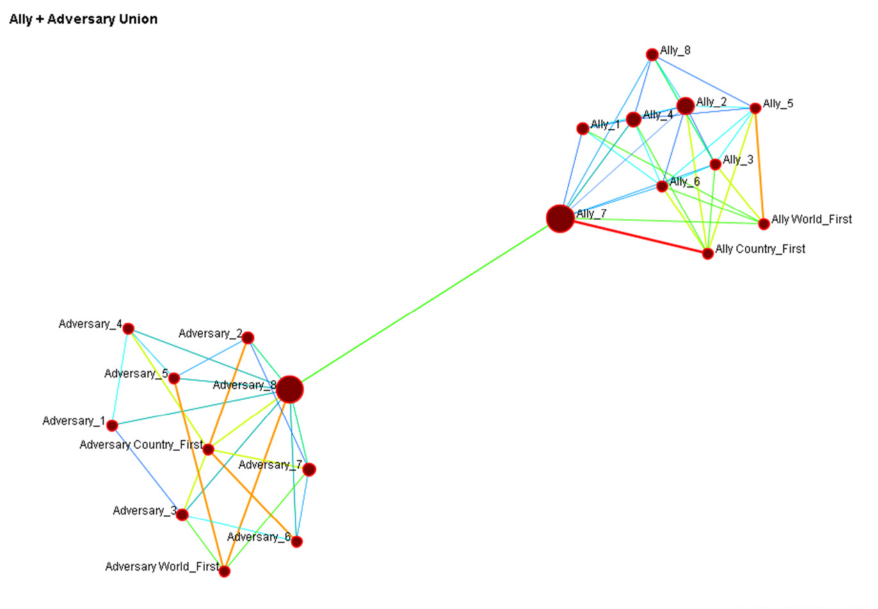


Figure 2. General Ally and Adversary Model with Anchoring Philosophical Beliefs (Nodes sized by Betweenness Centrality, Links colored by weight with red as highest weight)

CMU represented the threat and policy dimensions through a stylized combination of beliefs in the model. Each agent was given a biased but random distribution of knowledge bits in the simulation for each of six (6) core beliefs (see also Figure). There were three primary beliefs each agent held that aligned with the deterrence calculus: Threat—I believe my country feels threatened; Deter—I believe my country can deter [left vague but generally interpreted by SMEs as referring to military actions]; Punish—I believe my country can punish [left vague but generally interpreted in military and economic terms]. There were three (3) action-related, or secondary beliefs that we derived through a combination and overlapping of the knowledge that forms the basic beliefs: I believe my country should support the non-proliferation treaty; I believe my country should have nuclear weapons; I believe my country should pursue integrated missile defense. The overlapping boundaries and use of knowledge to support combinations of beliefs allow agents to potentially develop non-linear beliefs (e.g. an agent could feel threatened, not be confident of their ability to deter adversaries while remaining confident they could punish adversaries, and at the same time not desire nuclear weapons while wanting integrated missile defense). The relation of knowledge to beliefs is shown in Figure 3.

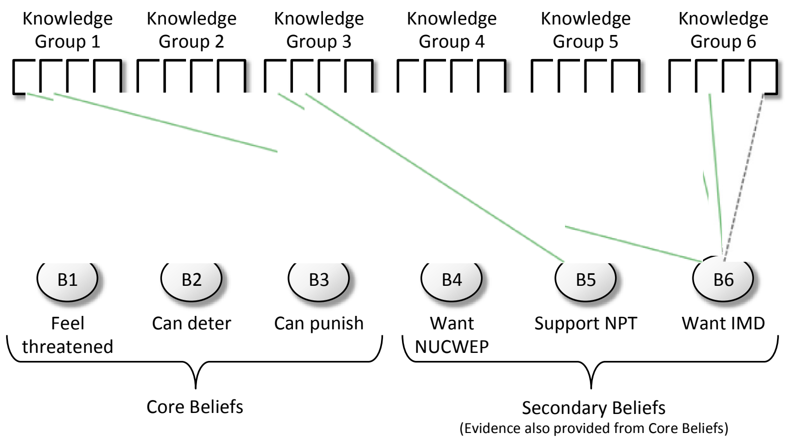


Figure 3. Modeling Core and Secondary Beliefs

Modelers represented the international future dimension, as well as policy dimensions, through the creation of links representing international ties between multi-agent entities. By default, the executives of the multi-agent entities/countries were always linked to other executives with weights of links to the instantiation of a region-specific model. Likewise, external affairs and economy and trade agents had international links by default; though, again, region-specific instantiation could result in deletion of those and other links (Figure omits these particular generic links).

Given that CMU's multi-agent simulation is an information and belief diffusion-based capability, modelers felt it appropriate to insert additional agents that would serve as the philosophical anchors of their respective nations' agents. In the generic model, both the adversary and ally were given two ideological agents, a "Country-First" perspective, and a "World-First" perspective (see also 2). For both agents, the Country-First perspective felt more threatened and was less confident of its ability to deter or punish. These anchors, for each country, were nearly impervious to exchange of knowledge with the potential for a change in belief, though we anticipated some future scenario-specific event that could alter these anchors (e.g. Pearl Harbor caused most isolationists to, at least temporarily, modify their professed beliefs). Examples of the effects of these ideological agents are shown below for both a generic ally, as well as a generic adversary.

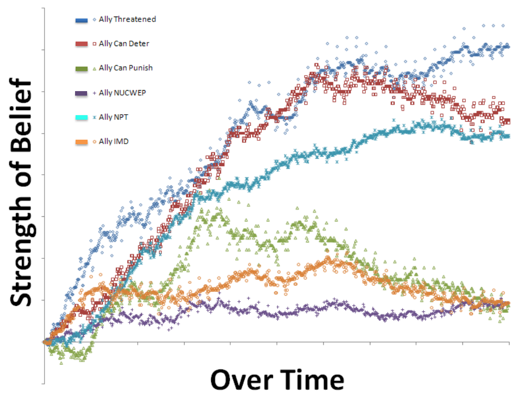


Figure 4. Allied distribution of philosophical knowledge over time

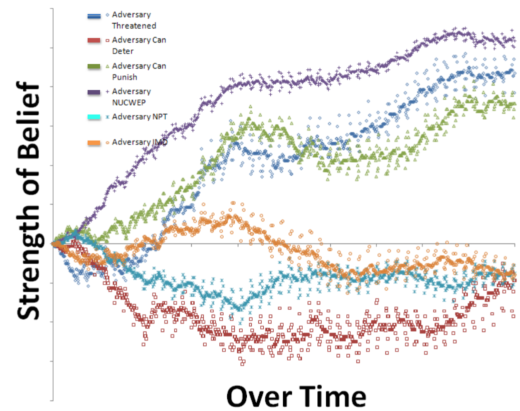


Figure 5. Adversary distribution of philosophical knowledge over time

INSTANTIATION OF A SPECIFIC MODEL

To create a model that represents a specific area of interest to the United States, the CMU instance of the model needs several key pieces of information, as well as intelligence estimates.

First and foremost, the target audience of the simulation must decide on the area of interest and the desired entities to represent. This research effort used Northeast Asia as its area of interest, with the addition of some peripheral entities, as illustrated below in Figure 8.

From this point, the analyst must:

1. For each entity, determine whether the entity can/should be reasonably represented as a monolithic actor. This is appropriate when the actor group suggested does not fit the eight stakeholder model well but the agent is still of interest (such as the IAEA or the UN), or if the expectation is that the country is only of peripheral interest and it makes sense to conserve modeling and SME effort.
2. For each core agent (one that is not represented as a monolithic actor), the customer, with SME assistance, must identify key philosophical anchors of that country (e.g. the Republic of Korea (ROK) could have some combination of anti-North Korea, pro-North Korea, and Business-First anchors). Figure 6 shows this representation for the United States, its eight (8) stakeholders and five (5) philosophical beliefs.
3. For each core agent, the SMEs will also need to assist in assigning, using the weights previously discussed, the interactions between the eight (8) agent types per country, as well as the international links between both monolithic and core agents. For this effort, CMU averaged the inputs of SMEs for international and anchor links, incorporated missing anchors as required, and modified the composition of agent types for one country by inserting a political party distinct from the other elements. Figure 7 depicts the interaction network between the USA and DPRK.
4. The SMEs and modelers have to assign distributions of initial beliefs to all actors—the three deterrence-calculus beliefs, as well as the distribution of the three derived beliefs.

5. For a specific scenario of interest, interventions and events should be defined in the context of their potential impacts upon specific stakeholders and what beliefs may be affected. The severity of an event is determined by which actors it impacts, how long it persists, and the amount of evidence the event provides for any particular belief.

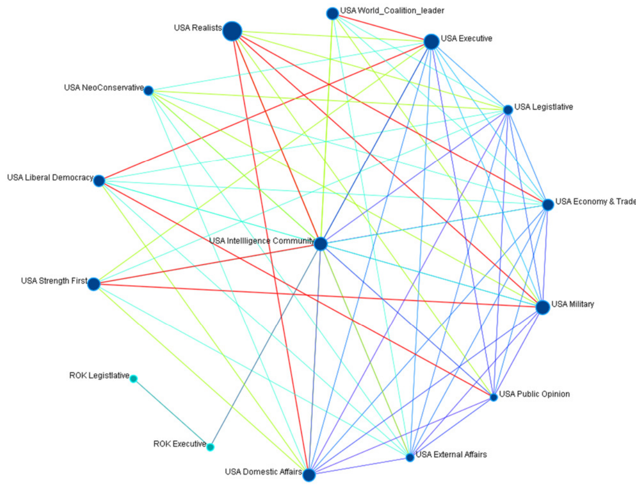


Figure 6. USA Internal Links between Stakeholders and Philosophical Beliefs. Links colored by weight, red highest and blue lowest. Nodes sized by Eigenvector value (node connected to well-connected nodes), and centered on betweenness centrality.

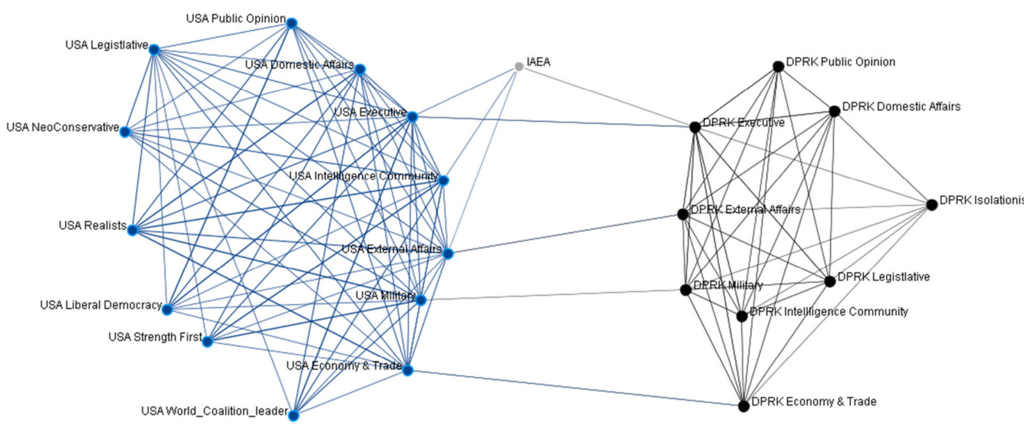


Figure 7. Depiction of USA and DPRK internal and external links between stakeholders and philosophical beliefs

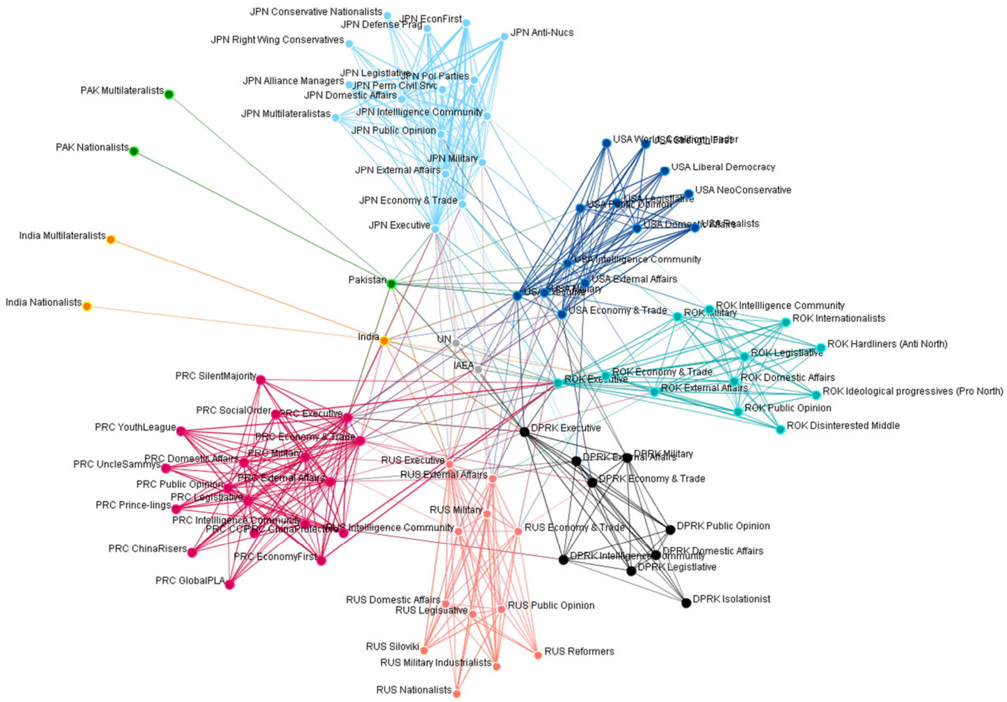


Figure 8. Stakeholder by stakeholder graphic. Uniform node sizes, color code by country, spring layout.

IMPACTS OF EVENTS AND INTERVENTIONS IN THE SIMULATION

A fundamental aspect to the CMU simulation capability is that agents' beliefs are influenced through the non-linear interaction of previously-held beliefs, perceptions of other's beliefs and knowledge; perceptions of own knowledge, perceptions of similarity between agents; seeking of new knowledge, as well as elements of randomness.

Events and interventions in this simulation are, therefore, stylized to ensure agents become aware of the event, that the awareness is reflected in the gaining, reinforcing, or contradicting of knowledge. Severity of events and interventions can be reflected in the amount of time the event remains present in the simulation, reflecting the amount of persistent conscious awareness of the event, and the various knowledge associated with the event. Severity could also be reflected in the amount of knowledge conveyed to any set of agents in any particular time period; the higher the number of knowledge bits transmitted per time period, the higher the probability that the knowledge will impact a belief(s) in a measurable way. By way of example, for the deployment of a dual-capable aircraft wing (i.e. capable of conventional and nuclear strike missions) into a region to have any deterrence or assurance effect, agents must become aware of the deployment with sufficient impact to affect their beliefs, while potentially overcoming their societies' philosophical anchors' influence. Another example could be the doubling of integrated missile defense (IMD) capabilities of Japan (JPN), Republic of Korea (ROK) and the USA. Figure 9**Error! Reference source not found.** depicts the simulations output of DPRK's six (6) beliefs of interest before becoming aware of such a change (the baseline condition). **Error! Reference source not found.**Figure depicts the set of DPRKs beliefs after the intervention (where the intervention occurs half-way through the simulation's run). Figure **Error! Reference source not found.** shows some not-surprising results: decrease in DPRK's perception of its ability to punish others and a perception that it loses deterrence capabilities. There is an upswing in the perception of feeling threatened, as well as a rise in the perceived desire to have nuclear weapons. Models like the base-case shown in Figure 9 are useful for confirming that the simulation has been coded to match the expectations of SME, allowing the exploration of counterfactuals, as shown in Figure 10. Note that, even in the base case, some of the relationships are not linear over time.

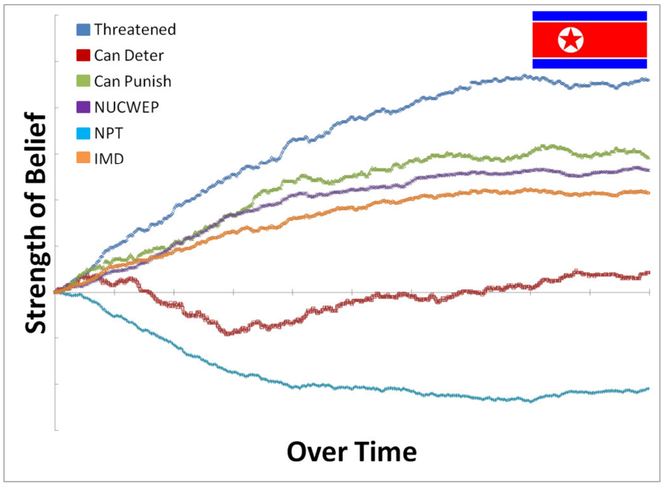


Figure 9. Baseline depiction of DPRK's 6 beliefs of interests

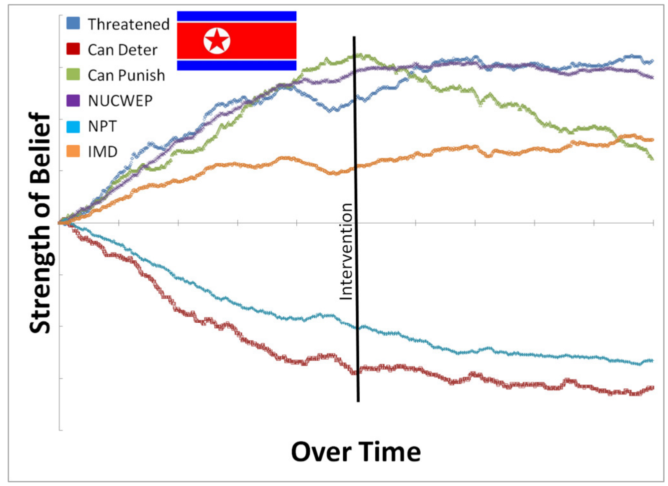


Figure 10. Depiction of DPRK's beliefs with one intervention, the doubling of JPN, ROK, and USA IMD capabilities

Finally, the interaction between Pythia and Construct has to, currently, be facilitated by human-to-human interactions. Pythia provides the probabilities and probability chains for discrete events in any given scenario or region of interest. Construct provides the changes in beliefs and perceptions driven by the probabilistic awareness of those events. Construct's changes then feed back into Pythia to assist in calculations moving toward the goal nodes representing the goals of the Nuclear Proliferation Review (NPR).

CONTRIBUTION TO THE UNDERSTANDING OF NUCLEAR STRATEGY

Devising nuclear strategy, of necessity, is a multi-dimensional problem for any nation. It is possible to view development of such strategy as an exercise in assembling a series of linear if-then problems (e.g. I feel threatened; therefore, I will build nuclear weapons), and seeking some optimal solution set (assuming such exists). It is more likely than not, however, that there are contradictions, or at least sets of mutually exclusive and non-exclusive competing interests, in virtually every nation's political calculus with respect to its nuclear strategy.

This model is a limited multi-dimensional effort at capturing some of the complexities of country's internal political processes, external political processes, as well as internal and external political dynamics. The model supports the cognitive truism that perceptions matter through the agents' ability to have perceptions of their own worldview, as well as perceptions of others they interact with. Though the perceptions are not full of nuance, their existence in the simulation supports the execution of actions and the holding of beliefs at odds with what could happen with an omniscient view of the world. It is feasible to increase the number of dimensions in this model, with the rising risk of an inability for clients/customers to understand the model, as well as a rising risk of loss of confidence in the outputs of the model's simulations.

Incorporation of nation's perceptions of themselves, their perceptions of others, their possession and perception of facts and knowledge that feed core beliefs, as well as tendencies of nation states to have national narratives are starting points for development of ways to implement our nation's nuclear strategy. Every force posture action, given exposure to allies and adversaries, with varying degrees of uncertainty, will contribute to perceptions and actions of others around the globe. Use of simulations to provide ranges of possible outcomes can help mitigate potential trouble for our own actions. Additionally, use of simulation to help identify potential trend lines and mitigation actions can help defuse situations before they become crises, as well as help modify and improve the model itself in the form of feedback loops.

It is clear from **Error! Reference source not found.** that there are interactions occurring between the six beliefs of interest. **Error! Reference source not found.**'s sharp drop in deterrence capability perceptions is worrisome. With a perception of low deterrence capabilities, people and nations can tend toward what others perceive as extreme reactions when confronted with situations they do not like; evidence of this can be found in the recent sinking of a ROK navy vessel and artillery shelling of an ROK island. Both events can be perceived as messages to the ROK that DPRK's military might should not be taken lightly, and that DPRK does indeed have an ability to punish, from which the ROK should be deterred from doing anything to provoke the DPRK.

This work contributes to the elements of nuclear strategy by incorporating not only the 5D model, but also the vortex model. The model deliberately incorporates the perceptions of international relations, has room for a multitude of DIME-related actions, and supports the propagation of information and beliefs about current events, while also supporting the less-prone-to-change national narratives for multiple countries and entities. By incorporating a multitude of dimensions, while staying clear of very high-dimensional visualization/understanding problems, this model allows the US to see the impact of actions related to nuclear strategy, as well as other, potentially competing, national priorities.

CAN THESE FINDINGS BE GENERALIZED?

The model, as it is built, is very generalizable within the realm of nuclear deterrence. It can be modified to address any set of regional, or non-regional countries of interest. The model can grow in complexity by incorporating a wider variety of competing policies (knowledge that supports or detracts from either the listed beliefs of interest or some alternative set of beliefs), as well as support customers'/clients' exploration of the input space. That exploration could reveal areas of absolute avoidance, as well as, potentially, numerous areas of desirable outcomes. Through multi-modeling and use of specific force posture actions in Pythia and its use of Time-Influenced Bayesian networks, modelers can predict reactions and consequences at both micro and macro levels.

Networks	**Node Types**								
Node Types	***Agent***	***Knowledge***	***Resource***	***Task***	***Event***	***Organization***	***Location***	***Role***	***Belief***
Agent	Social *“Who knows who”*	Knowledge *“Who knows what”*	Capabilities *“Who has what”*	Assignment *“Who does what”*	Attendance *“Who attends what”*	Membership *“Who belongs to what org”*	Agent Location *“Who is where”*	Role *“Who has what roles”*	Belief *“Who believes what”*
Knowledge		Information *“What informs what”*	Training *“What resources are needed for training”*	Knowledge Requirements *“What knowledge is task critical”*	Education *“What event teaches what”*	Organizational Knowledge *“What org knows what”*	Knowledge Location *“Where is what learned”*	Role Requirements *“What must be known to perform what”*	Knowledge Influence *“What knowledge informs what?”*
Resource			Substitution *“What can replace what”*	Resource Requirements *“What tasks require what”*	Event Requirements *“What events require what”*	Organizational Capability *“What org can do what”*	Resource Location *“Where is what”*	Role Requirements *“Who needs what resource to do what ”*	Resource Beliefs *“What beliefs are required to use what”*
Task				Task Precedence *“What must happen before what”*	Event Agenda *“What tasks occur at what?”*	Organizational Assignment *“What org does what”*	Task Location *“Where is what done”*	Role Assignment *“What roles do what”*	Belief Requirements *“What beliefs require what tasks”*
Event					Event Precedence *“What events happen before what”*	Organizational Responsibility *“What org is putting on what”*	Event Location *“Where is what event”*	Role-Event Requirements *“What roles are often present at what”*	Belief Attendance *“What beliefs influence participation at what”*
Organization						Inter-Organization *“What org works with what”*	Organization Location *“Where is the organization”*	Organization Role *“What org has what roles”*	Organizational Culture *“What beliefs are common”*
Location							Proximity *“What is near what”*	Location Roles *“What roles are common where”*	Significant Locations *“Where is associated with what beliefs”*
Roles								Inter-Role *“Who knows what”*	Significant Roles *“What roles are associated with what beliefs”*
Beliefs									Belief Influence *“What beliefs influence what”*

Appendix 3 Meta-Network table of nodes and 45 networks – precise semantics will depend on text corpus (Lanham Morgan and Carley 2011). Derived from Diesner and Carley 2005 and Carl

SME ELICITATION FOR THE CANS EFFORT

POC: Gregory Giles, SAIC, gilesg@stratcom.mil; Elizabeth Witter, Innovative Analytics and Training, Elizabeth.Witter.ctr@js.pentagon.mil

MODELS AND SIMULATIONS

Modeling and simulation teams from Carnegie Mellon University (CMU) and George Mason University (GMU) refined existing computer models that enabled users to see the potential outcomes of nuclear deterrence-related interactions amongst nations. Each modeling team undertook independent efforts to access SMEs and use these insights to fine-tune their respective model designs. The two modeling teams also used a common data source, specifically, the crowd-sourcing report produced by Monitor 360. Finally, while the two models did not share files, SME inputs obtained for the CMU model were fed into the GMU model.

CMU: CONSTRUCT

To help tailor a Construct model to the CANS problem set, CMU used a written questionnaire to collect SME insights on actors and their relationships. Developed by CMU modelers and vetted by the CANS Theory Team beforehand, the questionnaire sought to build a concise profile of each key country in the model. The questionnaire addressed how each country believed it was influenced by the other, who key stakeholders were within each country when it comes to nuclear policymaking, and the belief systems reflected by ideological groupings in each country. The authors passed the completed questionnaires to the CMU team and arranged teleconferences for the modelers to discuss the questionnaires individually with each SME. This approach was taken to preserve divergent views that can be subsumed by group interactions where a single individual can dominate or where there may be a tendency to drive towards a "lowest common denominator."

CMU modelers walked the SMEs through the questions and invited them to explain their answers. This approach was important as some interviewees were uncomfortable assigning numerical values to qualitative attributes, such as the strength of a relationship between specific actors. Typically, as the SMEs explained such relationships in qualitative terms, they became more comfortable with the numerical values they had assigned. The interviews also benefitted the CMU modelers in that the SMEs were able to convey deeper meaning to the numerical values they assigned. This elaboration highlighted subtleties that, according to the CMU team, improved model design.

GMU: PYTHIA

The GMU team met with SMEs to develop and tailor a generic model for two of the CANS efforts games. The GMU team collected data from SMEs knowledgeable in specific regions. These inputs were used to develop the specific regional models for use in the USAF and USN-coordinated CANS games. GMU's Pythia model also used Monitor 360's crowd- sourcing report.

GAMES

CANS incorporated three distinct multi-player games as a tool for exploring and testing hypotheses about US nuclear weapons policy. The Office of the Director of National Intelligence (ODNI), United States Air Force (USAF), and United States Navy (USN) each conducted a game independently, but coordinated with CANS.

Games enable decision-makers and SMEs to test theories about a broad range of issues in a non-attribution environment. For CANS, scenarios and role-plays were especially helpful for developing insights on the complexities of nuclear strategy. Game participants were SMEs who contributed their knowledge of other countries/regions, policy, strategy, military capabilities and weapons to their assigned team's group discussions on nuclear strategy. Across all of the games, team members contributed their varied expertise and willingly explored differences of opinion.

The DNI, USAF, and USN games included significant amounts of group discussion, which is one method of SME elicitation. SMEs were divided into teams in ways that best supported the game objectives. The DNI game divided teams by country knowledge. For the USAF and USN games, parallel teams were similar in size and experience breakdown (e.g., weapon systems, intelligence, and operational planning). These group settings yielded much data and drew out multiple arguments on controversial issues. However, it is important to note potential drawbacks to this type of elicitation. If a particular individual or sub-group is allowed to dominate the group discussions, the group's overall effectiveness may suffer. Challenges for SME elicitation in group settings, such as a game, include drawing out alternative viewpoints, encouraging quieter team members to share their thoughts, and managing more dominant team members' contributions.

CROWDSOURCING

Crowdsourcing was used in the CANS project to develop insights from non-U.S. perspectives. The SMEs included academic and policy experts, former officials, thought leaders, and influencers who could speak authoritatively on a regional level, have traveled throughout the region, and could discuss the perspectives of more than one country. The Monitor 360 team used a three-stage approach for this effort. Team members conducted initial interviews and open-source research to draw out a range of opinions regarding deterrence, assurance, and proliferation in East Asia, particularly Japan, South Korea, and China. In the second stage, Monitor 360 conducted in-depth SME interviews and developed hypotheses about segments of beliefs and opinions present in each country. During the third stage, Monitor 360 conducted a series of interviews to validate and refine the hypothesized segments and develop underlying narratives that explain the perspectives detected in each country.

ACADEMIC OUTREACH

To further supplement its SME elicitation techniques, the CANS project commissioned research papers to initiate discussion and development of the types of analysis and techniques applicable to nuclear deterrence and assurance issues. The CANS team reached out to well-known and respected scholars from the social sciences (sociology, anthropology, economics) and the physical sciences (physics). In addition, the team sought emerging scholars who are well-positioned to add new perspectives to the discussion of nuclear strategy.