<Your Name>

Validation and Agent Based Models

Prof. Kathleen M. Carley

POC: Kathleen M. Carley (PI)
Wean 1323
ISR, SCS, Carnegie Mellon
5000 Forbes Ave.
Pittsburgh, PA 15213 USA

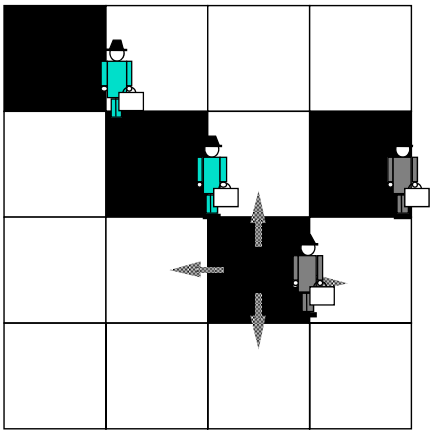
412-268-6016
kathleen.carley@cs.cmu.edu

Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

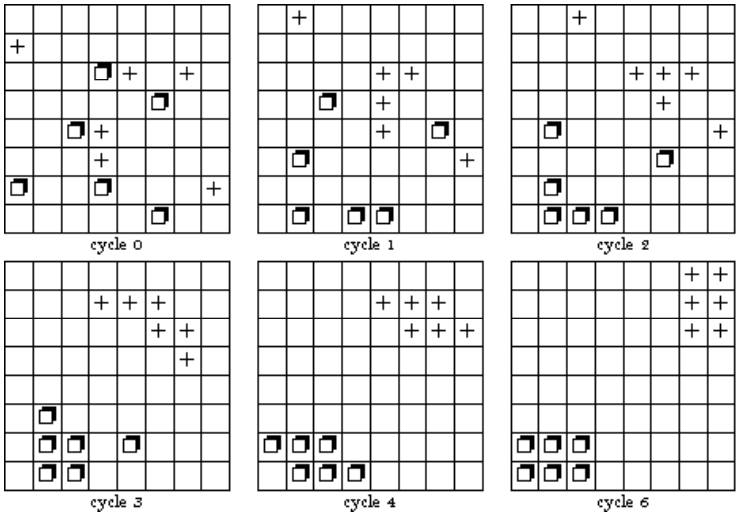
Segregation

I prefer to be near at least one other like me



<Your Name>

The Dynamics of Segregation
Sakoda 1971, P. 127



The Model …

- Input
  - Military Bases
  - Census data – social, economic, occupati
  - School district data
  - Worksite and entertain
  - Hospitals and clinics lo
  - Social Network charact
  - IT communication proc
  - Wind characteristics
  - Spatial layout
  - Disease models (symp
  - OTC and Prescription d
  - Attack or event
  - Interventions
- Illustrative Output
  - OTC & Prescription drug sales
  - Insurance claim reports (Dr. visits)
  - Emergency room reports
  - Absenteeism (school and work)
  - Infected, Contagious, Mortality

Agents move in networks which influence what they do, where, with whom, and what they know, what disease they get, when, how they respond m, etc. Major difference in netwo l disease nd age.

Vast quantities of data
Real and Virtual
Format of virtual to match real
But real data is …
Incomplete
Diverse sources
Inconsistent
Different levels of granularity

<Your Name>

Carnegie Mellon
ISR Institute for Software Research

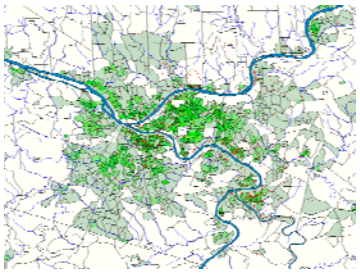
BioWar – City Level Impact of Epidemiological Events

City Description Manager

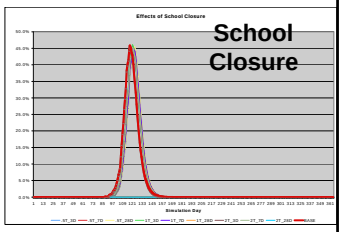
Disease Profile Manager

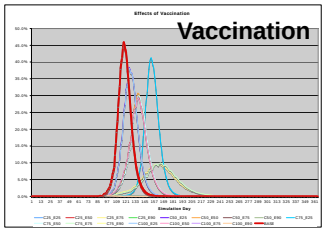
Attack Scenario Manager

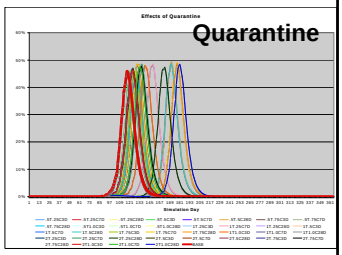
Output Manager



Drug Purchases







CASOS

Carnegie Mellon
ISR Institute for Software Research

Types of Models

- Intellective -v- emulation
- Stochastic -v- deterministic
  - Stochastic use random number generator (Monte Carlo)
  - Special case of deterministic is chaotic
- Steady State -v- Dynamic
  - SS use equations defining fixed relations
  - Dynamic relations change in response to signal
- Continuous -v- Discrete
  - Continuous periodically all equations are solved and state updated
  - Discrete maintains a queue of events and only items related to queue are solved
- Rule based -v- equation
- Learning -v- static -v- optimization
- Centralized -v- distributed (multi-agent)
  - One controller
  - Many individual controllers
- Local -v- distributed
  - Local – one machine
  - Distributed - runs on a network of machines
- System Dynamic -v- multi-agent -v- multi-agent-network

CASOS

CASOS

<Your Name>

How to Use Computational Models

- Test bed for new ideas
- Predict impact of technology or policy
- Develop theory
- Determine necessity of a posited mechanism
- Decision making aids
- Forecast future directions
- What if training tools
- Suggest critical experiments
- Suggest critical items for surveys
- Suggest relative impact of different variables (factors)
- Suggest limits to statistical tests for non-linear systems
- Substitute for person, group, tool, etc.. in an experiment
- Hypotheses generators

Definitions

- Validation – a set of techniques for determining whether or not a model matches with other models and/or matches with non-computational data. An extreme form is defining validity as a computational model's predictions match those from a set of non-computational data.
- Calibration – a set of techniques for tuning a model to fit detailed non-computational data.
- Training – procedures for supplying data and feedback to computational models that can learn.
- Verification – a set of techniques for determining whether the model behaves as intended. Ensures that the specification is complete and that mistakes have not been made in implementing the model. Also referred to as internal validity.
- Docking – a set of techniques for determining the level of comparability or equivalence of two models

<Your Name>

Carnegie Mellon
ISR

Verification

- No computational model is ever be fully verified
  - You can't guarantee 100% error-free implementation
- A high degree of statistical certainty is all that you can expect
- Statistical certainty increases with testing

Carnegie Mellon
ISR

There is a Science of Validation

- Developed for physical systems
- Integrated with analysis

1. Have a developed model
2. Have a set data that covers the IO space
3. Design a virtual experiment to cover the IO space
4. Conduct a virtual experiment
5. Build the response surface model
6. Compare the real and virtual data

- Issues are many
- Social systems are different

1. Model never fully developed
2. Clean accurate complete data is often not possible
3. You can design but can you analyze?
4. Insufficient computational and storage space
5. Tool large to build
6. SAS isn't robust enough!

<Your Name>

Carnegie Mellon
ISR

Why is validation difficult?

- Science of validation
  - Is based on physical systems
  - Socio-technical models have different characteristics that make validation infeasible
  - Assumptions are not true of complex socio-technical socio-cultural socio-psychological systems
- Characteristics
  - Large number of variables – so it is infeasible to construct entire response surface
  - Strongly complex
- Assumptions that are violated
  - Accuracy of empirical data
  - Continuity – but social systems have discontinuities
  - No covariance – but social systems have interacting and co-varying variables with no-apriori way of choosing between them
  - Consistency – but social systems have temporal variations in the relation of variables to each other

CASOS

Carnegie Mellon
ISR

Validation

- No computational model of a complex human socio-cultural system is ever fully validated
  - You can't guarantee 100% predictive accuracy
- The purpose of validation is increasing usefulness not proving accuracy
- Core validation questions:
  - Does the model address the right questions
  - Does the model provide accurate information about the system being modeled
  - Will this make the model actually used

CASOS

CASOS

<Your Name>

Methods and Levels for Validating a Computational Model

- Validation techniques vary in
  - Method
  - Level
  - Intensity
  - Purpose
- Similar approaches can be used for
  - Calibration
  - Training
  - Verification
  - Docking

Validation Levels

- Internal validity - error free code (verification)
- Parameter validity - parameters match
- Agent validity – agents match
- Process validity - processes fits
- Face validity - right type of things
- Pattern validity - pattern matches observed
- Requirements validity – requirements make sense
- Value validity - values match
- Theoretical validity - theory fits

<Your Name>

Carnegie Mellon
ISR

A Caveat

- Computational modeling is sufficiently complex that a single individual in a single research period (e.G. 6 months to a year) can not build, analyze, and validate a computational model.
- Most models take multi-person years to build and analyze.
- Data collection and analysis from a virtual experiment often takes as long as a human experiment and requires statistical training comparable to that required for human experiments.

CASOS

Carnegie Mellon
ISR

Validation!

- Validation is often inappropriate and impossible for complex socio-cultural models
- KISS models – are only meant to demonstrate a principle – you can't validate
- Veridical models – are so detailed that data doesn't exist to completely validate
- Validation can be done in parts
  - Input – parameters, starting conditions match the real world
  - Process – mechanisms have been observed in real world
  - Prediction – outcomes are observed in real world

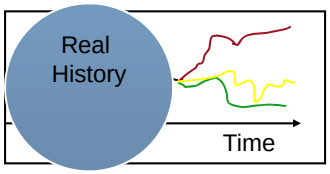
CASOS

CASOS

<Your Name>

Validation to Historical Point

- Validation to a historical point
  - Leads to over-fitting
  - Assumes the world is static
  - Ignores tipping point phenomena



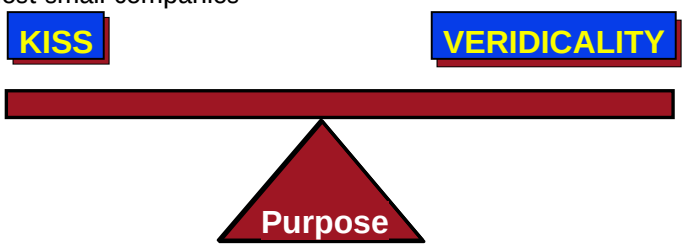
- This is the approach used in many "statistical" models
  - Data mining
  - Econometrics
  - OK for short term – 6 months – 1 year (ok for marketing)
- Real world
  - Many punctuation points
  - Tipping points have been observed

Agent Based Models provide the "Space of possibilities"

Balance

Michigan
Santa Fe
Brookings
Sandia
GMU
Most small companies

CMU
U Penn
Los Alamos
UVA



<Your Name>

Carnegie Mellon
ISR

Face Validity

- Is the model a reasonable simplification of reality?
- Techniques to increase face validity:
  - Set parameters based on real data
  - Model a specific organizational or inter-organizational process
  - Show that others have made similar assumptions
  - Discuss model limits and how left out factors may or may not affect results
  - Don't over-claim model applicability

Carnegie Mellon
ISR

Illustration of Face Validity

MODEL	REALITY
Simulated Annealing	Organizational Strategic Adaptation
system	organization's CEO or central unit
state	organizational design
current state	current organizational design
temperature	risk aversion
accepting a cost	taking a risk
increasing move	
high temperature means	liability of newness
accepting many cost	
increasing moves	
move set	re-design strategies
heuristic optimization process	satisficing & BR process
minimize cost	maximize performance
cooling schedule	approach to becoming risk averse
proposed state	proposed new design
evaluation of proposed state	limited lookahead, anticipation of future
state evaluation	observed performance

<Your Name>

Carnegie Mellon
ISR Institute for Software Research

(Social) Turing Test

The model does the task it seeks to explain.
1) substitutability.
2) Turing test.
3) social Turing test.

Construct a collection of social agents according to the hypotheses and put them in a social situation, as defined by the hypotheses. Then recognizably social behavior should result.

Aspects not specified by the hypotheses, of which there will be many, can be determined at will.

The behavior of the system can vary widely with such specification, but it should remain recognizably social.

Carnegie Mellon

Cognitive Architecture

Knowledge

Increasingly Rich Situation

Increasingly Limited Capabilities	Nonsocial Task	Multiple Agents	Real Interaction	Social Structural	Social Goals	Cultural Historical
Omniscient Agent	goal directed produces goods uses tools uses language	models of others turn taking exchange	face-to-face	class differences	organizational goals	historically situated
Rational Agent	reasons acquires learns	learns from others education negotiation	mis-communication	promotion social mobility	competition cooperation social cognition	emergent norms *Seas* *Economic Models* *Cultural Transmission*
Boundedly Rational Agent	satisfices task planning adaptation	group making	social planning coercion priority disputes	altruism uses networks for information boundary spanners *Garbage Can Model* *Sugarscape, AAIS*	delays gratification moral obligation *RTE* *VDT* *TAEMS*	gate keeping role emergence *CORP, HITOP-A, ACTION, ORGAHEAD, Organizational Consultant*
Cognitive Agent	compulsiveness lack of awareness multi-tasking	group think	spontaneous exchange social interactions *Soar*	automatic response to status cues *Construct*	group conflict power struggles	develop language institutional change
Emotional Cognitive Agent	habituation variable performance	protesting trust *Athena*	play rapid emotional responce cons	campaining	team player	norm maintenance ritual maintenace advertising **MODEL SOCIAL AGENT**

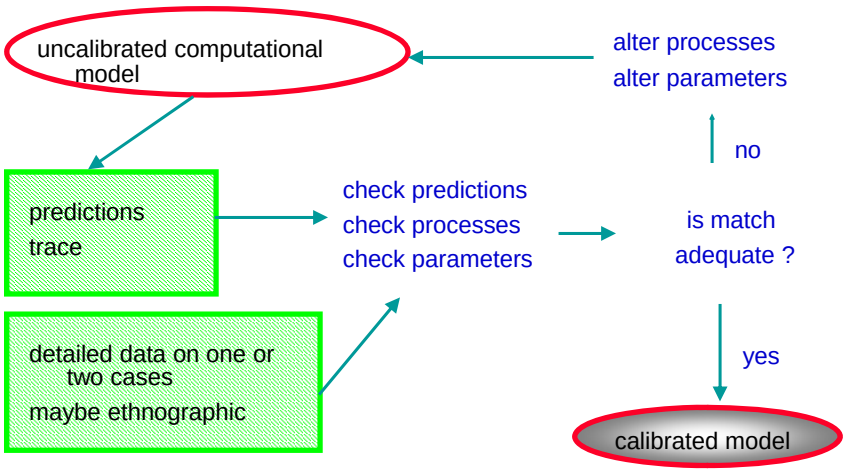
<Your Name>

Calibration

- Calibration involves fitting a computational model to a set of data
  - May require programming (adding modules or new processes)
  - May require parameter setting
- Have available detailed data on one or more cases
- Calibration is often the only validation step carried out for emulations
- Calibration demonstrates that model can match non-computational data

Iterative build-check process

Calibration Cont.



<Your Name>

Carnegie Mellon
ISR

Locating Cases

- Ideally:
  - Use a set of cases that span the key categories you are concerned with
- Next best:
  - Choose 2-4 cases that represent typical behavior and 1 or 2 that represent atypical behavior
- In practice:
  - Most intellective models are not calibrated
  - Lucky to have even one case with sufficient detail
  - Often detailed case is a matter of opportunity
- Sources:
  - Archival data, ethnographies, participant observation

CASOS

Carnegie Mellon
ISR

Multi-expert Problem

- What if experts or cases disagree?
- Typical solution: average the two cases
- Alternative: put in both cases as options with a certain probability of occurring
  - Probability:

  Equally weighted

  Weight can reflect degree of agreement

NOTE: For rule based models detailed cases may be the opinions of experts

CASOS

CASOS

<Your Name>

Simple Validation

- Simple techniques for seeing if model results are reasonable.
- Techniques to demonstrate validity:
  - Are there stereotypical facts about the problem that this model generates; E.G., Models of organizational evolution should predict liability of newness.
  - Are there behaviors that any model of this ilk should generate; E.G., All diffusion models should generate an s-shaped adoption curve, all neural networks should take a long time to train.
- These are non-surprising findings but if model can't generate them then it is not valid.

General Validation

- Validation involves testing a computational model's predictions given a set of non-computational data
- Have available the results of a virtual experiment
- Have available non-computational results
  - May be archival, survey, experimental
- Validation is sometimes done on uncalibrated models
- Validation demonstrates that model's predictions match non-computational data

<Your Name>

Characteristics of Validation

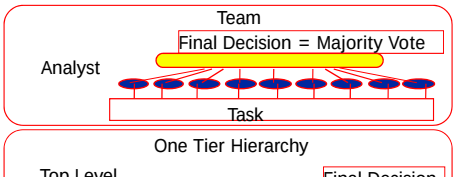
- Level of validation:
  - Pattern - same trends are observed
  - Value - same values are observed
- For multi-agent models:
  - Group or organizational level - matches overall behavior of collection of agents
  - Agent level - matches specific entities behavior
- For stochastic models:
  - Point - on average behavior is the same
  - Distribution - distribution of results is the same
  - Detail match - one entire run is the same

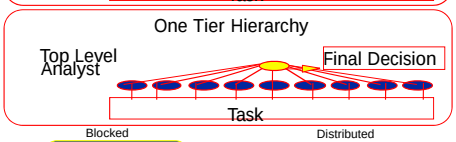
Data for Validation

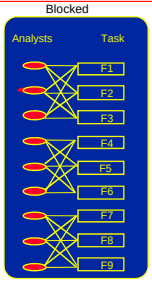
- Type: anything
  - May be archival, survey, experimental
- Quantity: high
  - Sufficient for statistical analysis
- Level of detail: low
  - Do not need the same level of process data that is needed for calibration

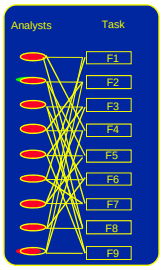
<Your Name>

Corp
(a predecessor to ORGAHEAD)

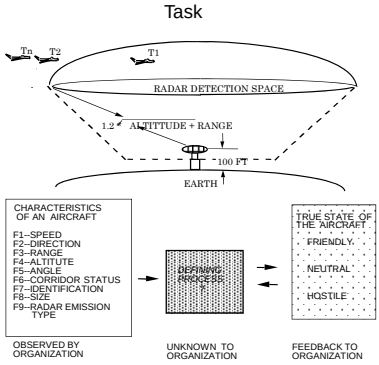








Used a subset that matched human experiment



Matched Analysis

Simulation

- Vary organizational design
- Vary task environment
- Measure performance as accuracy
- Monte Carlo 19683 cases
- Estimate of performance on average

Corporate Data

- Vary organizational design
- Vary task environment
- Measure performance as actual/potential severity
- General performance
- 69 cases, technological disasters

Matched Set

Predict performance

What if analysis: if organization did/did not shift what would be impact

<Your Name>

Carnegie Mellon ISR

Archival Match

Training	Prediction: Performance in General	Prediction: Performance During Crisis
Experiential	2.10(21,0.18)	2.38(21,0.11)
Operational	1.83(48,0.10)	1.42(48,0.07)

Training	Observation: Performance in General	Observation: Performance During Crisis
Experiential	1.86(21,0.17)	2.38(21,0.13)
Operational	1.83(48,0.09)	1.46(48,0.08)

Note: Number of cases and standard errors are in parentheses

Carnegie Mellon ISR

Sub-Optimalities and Performance

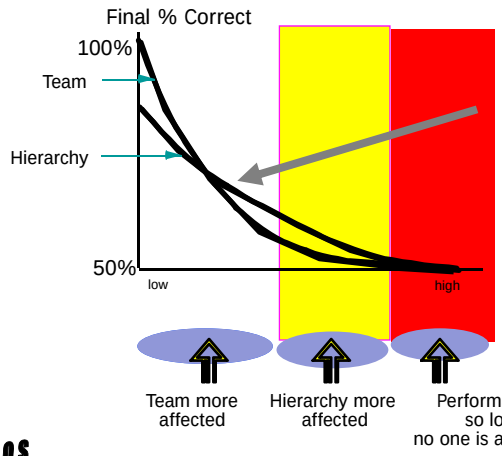
Training	Prediction: 1	Prediction: 2	Prediction: 3
Experiential	2.67(9,0.17)	2.18(11,0.12)	2.00(1,0.00)
Operational	1.21(24,0.15)	1.68(22,0.10)	1.00(2,0.50)

Training	Observation: 1	Observation: 2	Observation: 3
Experiential	2.56(9,0.18)	2.27(11,0.20)	2.00(1,0.00)
Operational	1.42(24,0.15)	1.50(22,0.11)	1.50(2,0.50)

Note: Number of cases, and standard errors are in parentheses

Carnegie Mellon
ISR

Hierarchies More Robust



Slope of curves, intercepts and hence the crossover point depends on the level of turnover, personnel experience, task complexity, interaction of errors ...

Carnegie Mellon
ISR

Reality: Teams Better

	TEAM		HIERARCHY		SEGREG		NON-SEG	
	exp	sop	exp	sop	exp	sop	exp	sop
Simulated	3.00	1.50	2.35	1.41	2.30	1.40	2.45	1.60
Human	3.00	1.50	2.35	1.46	2.10	1.42	2.64	1.80
	(1)	(4)	(20)	(44)	(10)	(43)	(11)	(5)

<Your Name>

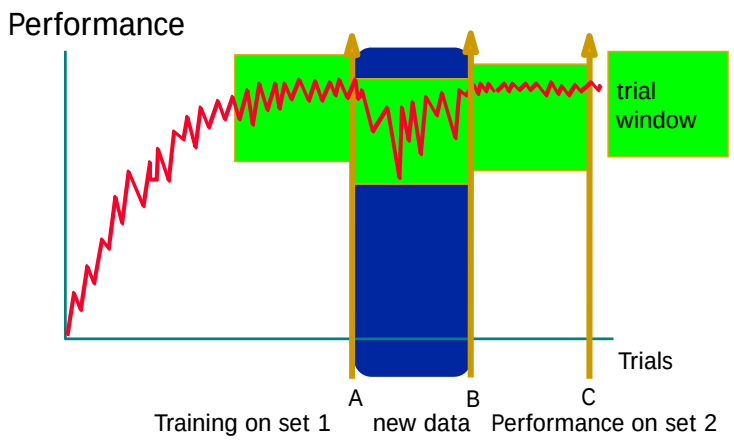
Carnegie Mellon
ISR

Model Training

- Use this when:
  - You have models that learn
  - You want to test the "goodness" of what they have learned
- Approach
  - Divide non-computational data into two sets
  - Train the model on first set
  - Generate model predictions to second set
  - Test model against second set

Carnegie Mellon
ISR

Impact of Training

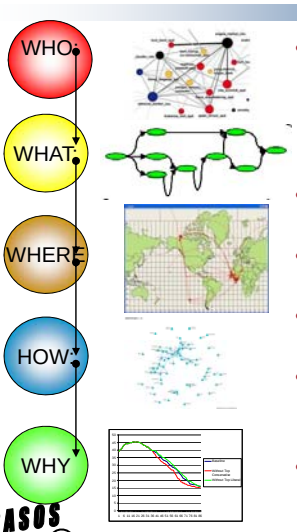


<Your Name>

Key Issues Impacting Validation

- Should the model be validated?
  - Intellective are for proof of concept – NO VALIDATION IS NEEDED
- Unequivocally measurable variables
  - If they don't exist don't validate – inducing errors of inference
- Veridicality of model
  - Greater veridicality greater empirical versus face validation
- Sophistication of Model Human Agent
  - Greater sophistication greater empirical validation at micro level, fewer agents and, CURRENTLY, less empirical validation at macro level
- Network focus
  - Many available data sets
  - Enhances re-use

Networks are NOT just About People

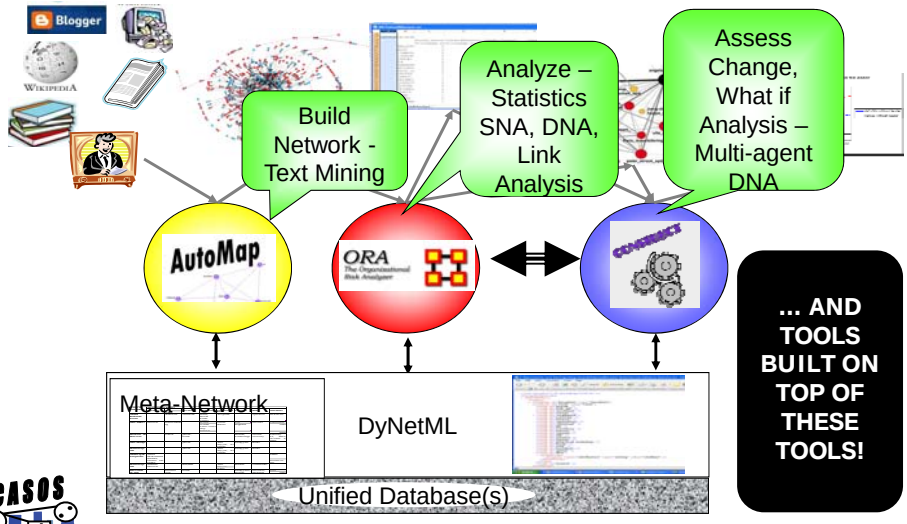


- WHO
  - Social network
  - Tribal network
  - Organizational network
  - Country alliance network
- WHAT
  - gant chart – what occurs before what
- WHERE
  - maps – sharing boundaries
- HOW
  - semantic networks – ideas to ideas
- WHY
  - belief network – cascades of beliefs
- **& Networks interconnect**
  - Activity network – who is doing what
  - Capability network – who has what resources

AND WHEN! TRAILS!

<Your Name>

Addressing these missions requires a suite of tools



Needed: A new science of validation

- Really should not even be talking about validation
- For most models only face validation is needed
- Validation is a team effort
- Need approaches that support
  - Prediction as evincing the space of possibilities
  - Inconsistency in effects
  - Discontinuities
  - Co-variance
  - Messy data

<Your Name>

Carnegie Mellon
ISR Institute for Software Research

References

- 2008, Behavioral Modeling and Simulation: From Individuals to Societies, Washington, DC: National Academy Press.
- 2006, Defense Modeling, Simulation, and Analysis: Meeting the Challenge, National Research Council, National Academy Press. Washington, D.C.
- AIAA. 1998. AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. American Institute of Aeronautics and Astronautics.
- Axelrod, R. "Advancing the Art of Simulation in the Social Sciences" *Simulating Social Phenomena. Conte, R., R. Hegselmann and P. Terna, eds. Springer: Berlin.*
- Axtell, R., R. Axelrod, J. Epstein and M. D. Cohen. 1996. "Aligning Simulation Models: A Case Study and Results" *Computational and Mathematical Organization Theory 1:123-141.*
- Bankes, S. C. 2002. "Tools and Techniques for Developing Policies for Complex and Uncertain Systems" *Proc. National Academy of Sciences 99(90003):7263-7266* http://www.pnas.org/cgi/ content/abstract/99/suppl_3/7263.
- Burton, R. "Validating and Docking: An Overview, Summary, and Challenge" *Simulating Organizations: Computational Models of Institutions and Groups. Prietula, M. J., K. M. Carley and L.* Gasser, eds. MIT Press: Cambridge, MA.
- Carley, K. M. 2002. "Simulating Society: The Tension between Transparency and Veridicality" Agent 2002 Conference on Social Agents: Ecology, Exchange, and Evolution. Macal, C. M. and D. L. Sallach, eds. Chicago, IL. Oct. 11-12, 2002. pp. 103-113.
- Carley, K.M. 2009, "Computational Modeling for Reasoning About the Social Behavior of Humans," Computational and Mathematical Organization Theory,15(1): 47-59.
- Edmonds, B. and D. Hales. 2003. "Replication, Replication and Replication: Some Hard Lessons from Model Alignment" Journal of Artificial Societies and Social Simulation 6(4) http:// jasss.soc.surrey.ac.uk/6/4/11.html.

CASOS

Carnegie Mellon
ISR Institute for Software Research

Extra

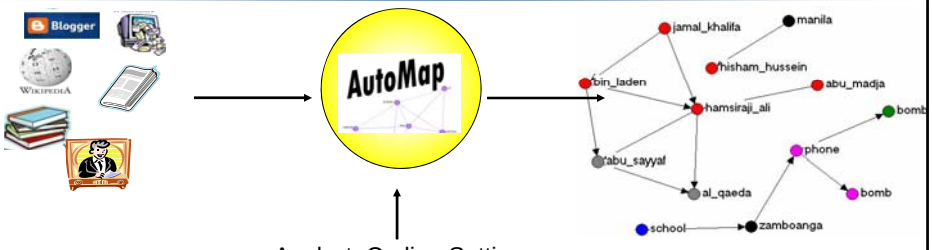
CASOS

CASOS

Refs

- Annino, J. S. and E. C. Russell. 1979. "The Ten Most Frequent Causes of Simulation Analysis Failure – and How to Avoid Them!" *Simulation 32(6):137-140.*
- Bankes, S. C. 1993. "Exploratory Modeling for Policy Analysis" *Operations Research 41(3):435-449.*
- Epstein JM, Axtell R. 1996. Growing Artificial Societies: Social Science from the Bottom Up. Cambridge, Mass.: MIT Press
- Gilbert N, Troitzsch KG. 1999. Simulation for the Social Scientist. Buckingham: Open University Press

AutoMap: Text Mining to Extract Networks



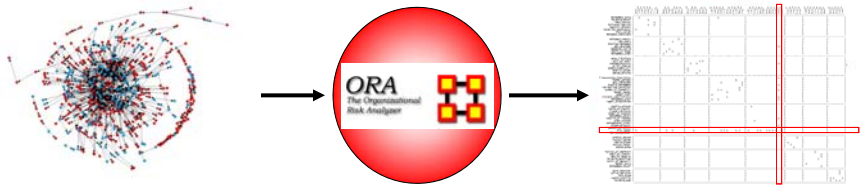
Analyst: Coding Settings

Core Capabilities
- Content Extraction (a.k.a. Content Analysis)
- Semantic Network Extraction (a.k.a. Mental Model Analysis)
- Meta-Network Extraction (a.k.a. Structural Analysis)
- Belief Extraction and Inference (a.k.a. Sentiment Analysis)

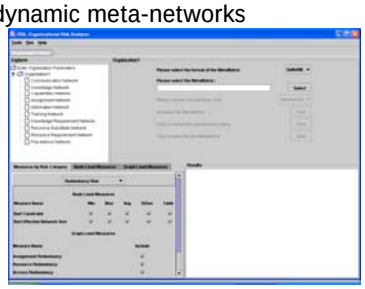
Illustrative low level component capabilities
- Anaphora resolution
- Deduplication
- Thesauri construction
- Stemming

<Your Name>

*ORA: SNA, DNA, and Link Analysis

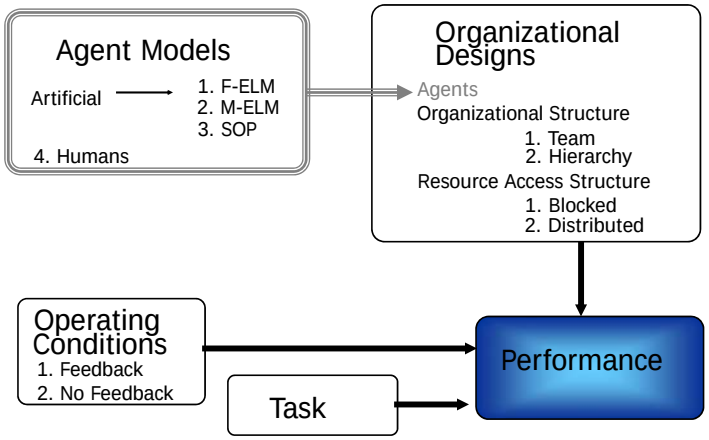


ORA: a DNA statistical analysis tool for locating patterns and identifying vulnerabilities in dynamic meta-networks



- Organized by function not measure
  - E.g., Key Entity Report
- Import/Export tools
- Linkage to mysql
- Visualization components
- Batch and thick-client
- Can handle large 10^6 networks quickly
- Supports data entry, editing, combining, cleaning, analysis, forecasting, visualization and storage

Illustration: CORP & Human Experiment



<Your Name>

Carnegie Mellon
ISR Institute for Software Research

Human Experiment

Two stages -

Stage 1) Analysts / Operators Decisions based on raw data

Stage 2) Commander/mid-level Decisions based on analyst's decisions

Each analyst sees 120 problems of form

l,m,h on 3 characteristics

speed is high, altitude is high, radar type is weapons emissions

set 1) 30 problems with characteristics a,b,c with feedback

set 2) 30 problems with characteristics a,b,c without feedback

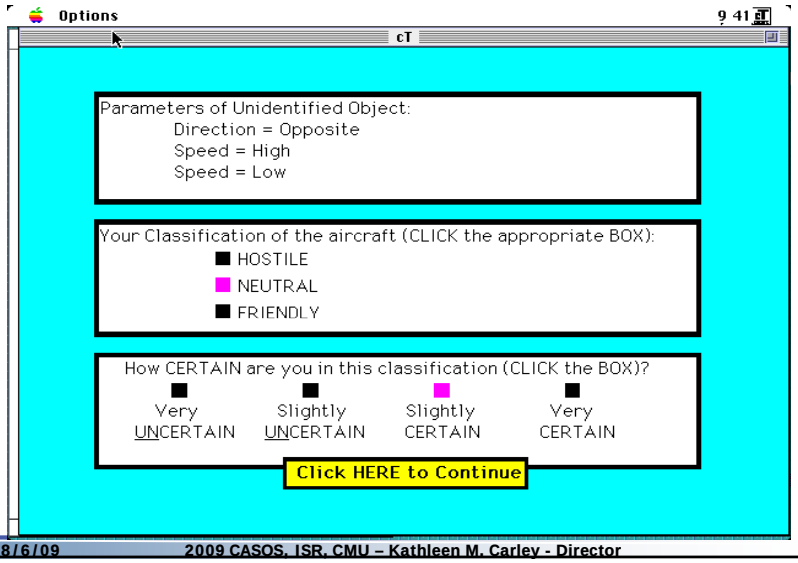
set 3) 30 problems with characteristics d,e,f with feedback

set 4) 30 problems with characteristics d,e,f without feedback and with equipment failures

CASOS

Carnegie Mellon
ISR Institute for Software Research

Details of Experiment

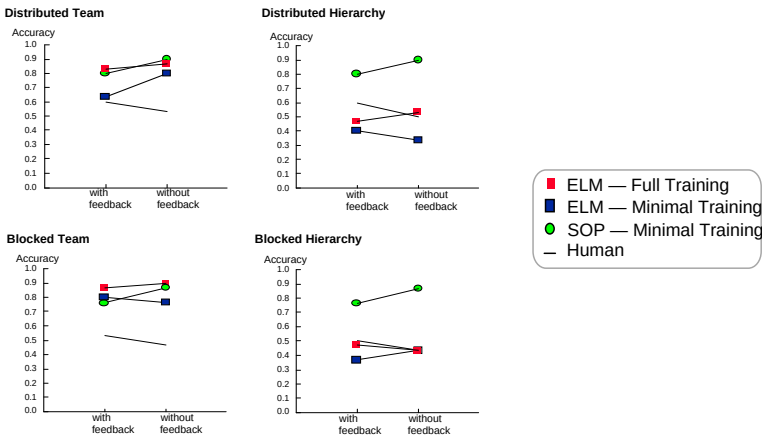


CASOS

CASOS

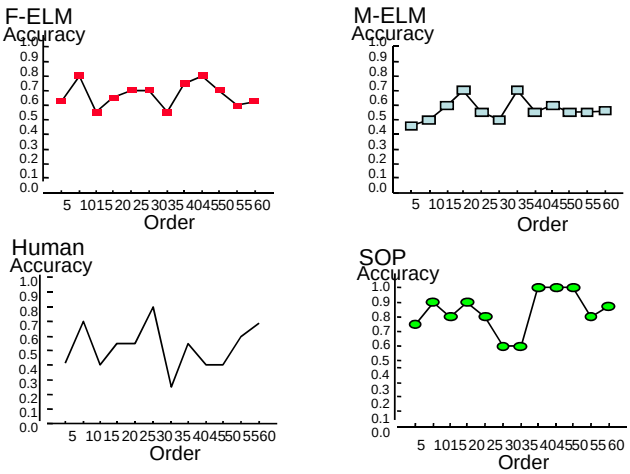
Carnegie Mellon
ISR Institute for Software Research

Verification by Pattern



Carnegie Mellon
ISR Institute for Software Research

Detailed Pattern



<Your Name>

Carnegie Mellon
ISR

Verification by Value

Organization's Accuracy by Agent Model and Organizational Design

Agent Training	Team		Hierarchy	
	Blocked	Distributed	Blocked	Distributed
ELM full	88.3%	85.0%	45.0%	50.0%
ELM min	78.3%	71.7%	40.0%	36.7%
SOP full/min	81.7%	85.0%	81.7%	85.0%
Human	50.0%	56.7%	46.7%	55.0%

Carnegie Mellon
ISR

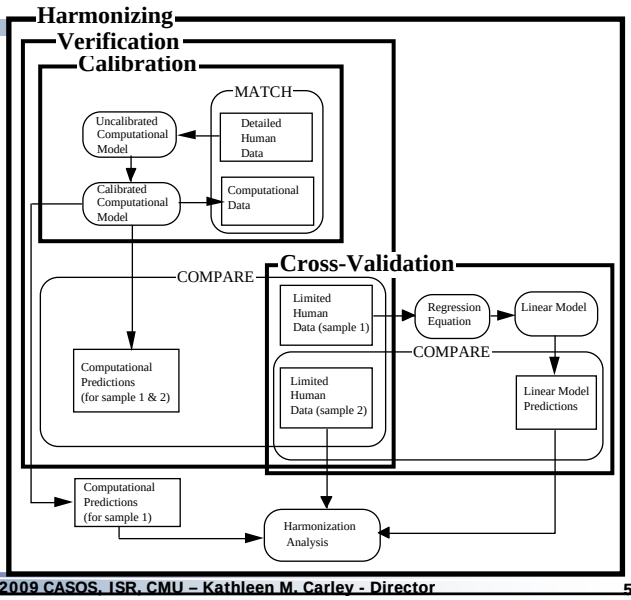
Harmonization

- Assessment of theoretical adequacy by comparison with a cross-validated linear model
- Harmonization involves contrasting the predictions of a computational model and a linear model
  - Requires enough cases that you have two samples large enough for statistical analysis
  - Requires that there is a reasonable linear model
- Harmonization can locate areas of the model where the embodied theory is inadequate

statistical comparison process

<Your Name>

Harmonizing

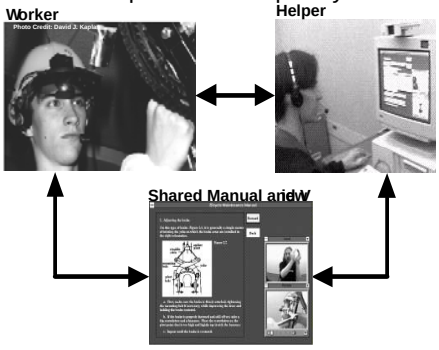


Locating the Linear Model

- Many sources for such a model:
  - Linear model of inputs
  - Easily collected data that might be used by management to make the same prediction the computational model is making
  - Model presented in the literature

Illustration & Harmonization

COMIT - impact of technology on task performance
Human data on a bicycle repair task using various technology (including vide an non-video)
Focus is on prediction frequency and order of communication actions



- Coded audio soundtracks
- Took half of subjects and created linear model and verified COMIT
- Linear model based on data accessible by managers
- Generated predictions for remaining subjects for COMIT and linear model
- Calculated R2 across subjects
- Data - 4 people, COMIT 2 people

COMIT Results

Action	Linear Model	COMIT	Correlation
Helper Acknowledgment	0.230	0.653	0.8083
Helper Request	0.498	0.141	0.7488
Helper Other	0.118	0.323	-0.3306
Worker Other	0.307	0.368	0.6886
Worker Question	0.215	0.011	0.4442
Worker Description	0.086	0.182	0.8148
Worker Acknowledgment	0.044	0	
Helper Help	0.310		

positive correlation means both models doing well/poorly in same areas, negative correlation means models tend to have opposite predictions