The Interaction Between Theory-Driven Inquiry and Data-Driven Discovery

Marshall Scott Poole
University of Illinois Urbana-Champaign

Some History

- Virtual Worlds Exploratory Project
  - Nosh Contractor, Northwestern
  - Jaideep Srivastava, Minnesota
  - Dmitri Williams, University of Southern California
  - Scott Poole, University of Illinois
  - Ron Burt, University of Chicago
  - 30 co-investigators (including Brooke Foucault, Yun Huang, Brian Keegan, Jeff Treem, Mengxiao Zhu)
- Call for White Papers on Future of SBE Inquiry

  "Data Driven Discovery in the Social, Behavioral and Economic Sciences"

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Data-Driven Discovery

- VWE Project: Game Data from MMORPG
  - Saved by Game Company for Transactional Purposes
  - Not Designed for Research
  - Typical—Multiple administrators, not much documentation
- Transform, Clean-up, Index
- Analytical Techniques
  - Extract Information
  - Identify Patterns
  - Machine Learning Predictive Models

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Data-Driven Discovery

- New Discoveries
  - Group Identification: Modular Networks of Dynamically Rearranging Groups
  - Prediction of Real World Characteristics from In-Game Behavior and Character Attributes
- Explananda: “Facts” to be Explored and Explained

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Theory-Driven Inquiry

- Traditional approach for SBE Sciences
- Theory: The Beacon of Inquiry
- Theory: Necessary as a Corrective for Poor Data
  - Sparse, costly data (at least relative to today's digital data avalanche)
  - Noisy, error-full data
- Theories of Data: The –metrics (psychometrics, sociometrics, econometrics, etc.)

Theory-Driven Inquiry

- The Theoretical Beacon as Lamp Post

- Data-Driven Discovery as Parking Lot Illuminator

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Q1. How Do We Think Theoretically in the World of Data Driven Discovery?

Data Driven Discovery as Surprise

- It Poses Puzzles and Predicaments

- Chance and Discovery

  “Chance favors the prepared mind”

  /ɫ°ɫɨ#ɓ˚ˡ–ˇ°ß

- Surprise as a Stimulus for Thought

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Making Sense of Surprise

- Deduction & Induction
- Pragmatists
  - Default mode is “habitual” action
  - We learn when we encounter problems/surprises/hitches
- Retroduction
  - Process by which we make sense of problems
  - Playful
  - Increasingly systematic

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Retroduction

Retroduction:

- “Moving back for a purpose”
- Stopped by a surprise, we move back to try to comprehend the surprise
- Fundamental move in knowing
- A key aspect: Abduction: “Leading away from”

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Abduction

- Proceeded by aesthetic apprehension of surprising thing
  - Unfettered delight and appreciation
  - Sense of satisfactory frame to adopt
- Abduction—more focused play
  - Entertain multiple theories/models/frames
  - Focus in on one

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

The Cycle of Retroduction

Recursive cycle:

1. “A surprising fact is noticed
2. An aesthetic (unfettered) exploration of qualities and relationships is made
3. Abductive reasoning is applied to make a guess that could explain the surprising fact
4. Deductive reasoning is applied to explicate the guess and ready it for testing
5. Once readied, inductive reasoning s applied to test and evaluate the guess
6. Abduction or deduction is used to respond to that evaluation (or new information is produced) and the cycle begins again until a hypothesis (or conditional purpose) has been fully engendered and is ready for formal explication and testing.”

P. Chiasson (2005). Abduction as an aspect of retroduction *Semiotica, 153*, 223-242 (quotation, p. 238; emphases mine).

Utility

- Recognition no one right process of thinking
- Balance among processes? Is there one? Should there be one?
- Sequencing of processes
- Methods for abduction?
- Algorithm???

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Q2. How Might Data Driven Discovery and Theory Driven Inquiry Interact During the Research Process?

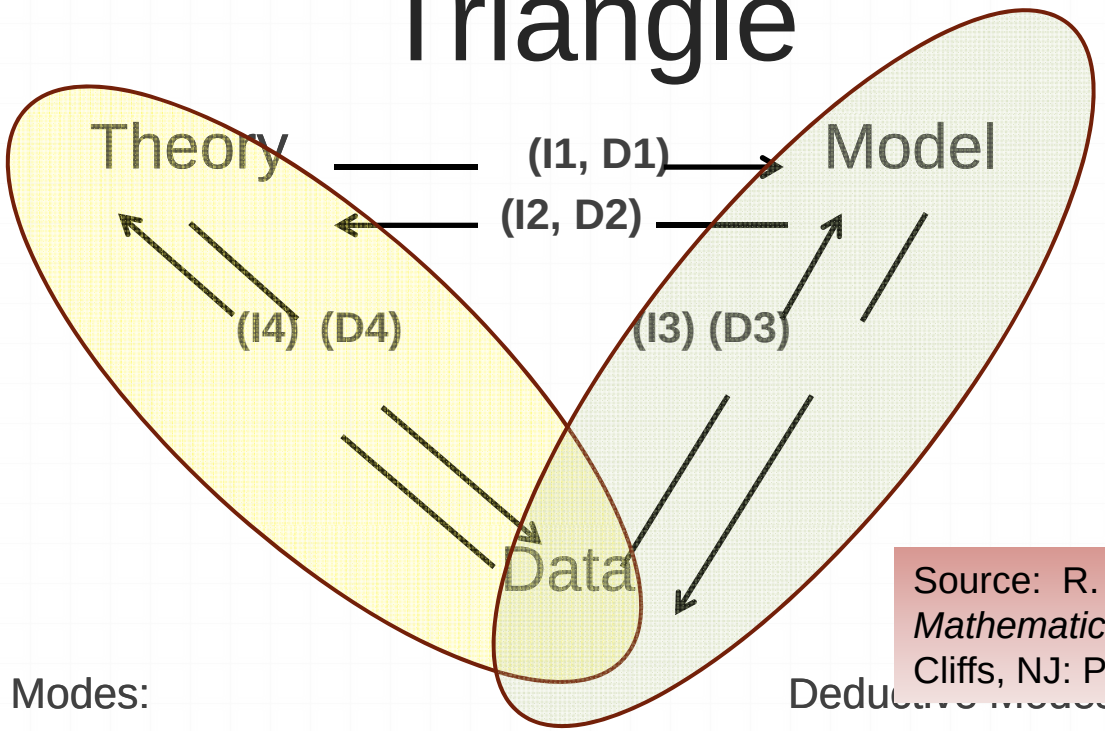
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Assumption

It's desirable to…

- Consciously think through the relationship of the elements of theory-driven and data-driven inquiry
- Bring both to bear in our scholarship

The Theory-Model-Data Triangle



A dialogue

Source: R. Leik & B. Meeker (1975). *Mathematical Sociology.* Englewood Cliffs, NJ: Prentice-Hall.

Inductive Modes:

I1. Mathematical generalization of theory

I2. Substantive interpretation of model

I3. Generalization of empirical results in model patterns

I4. Substantive interpretation of data

Deductive Modes:

D1. Formalization of theory

D2. Deriving hypotheses from model

D3. Model prediction of data

D4. Substantive prediction

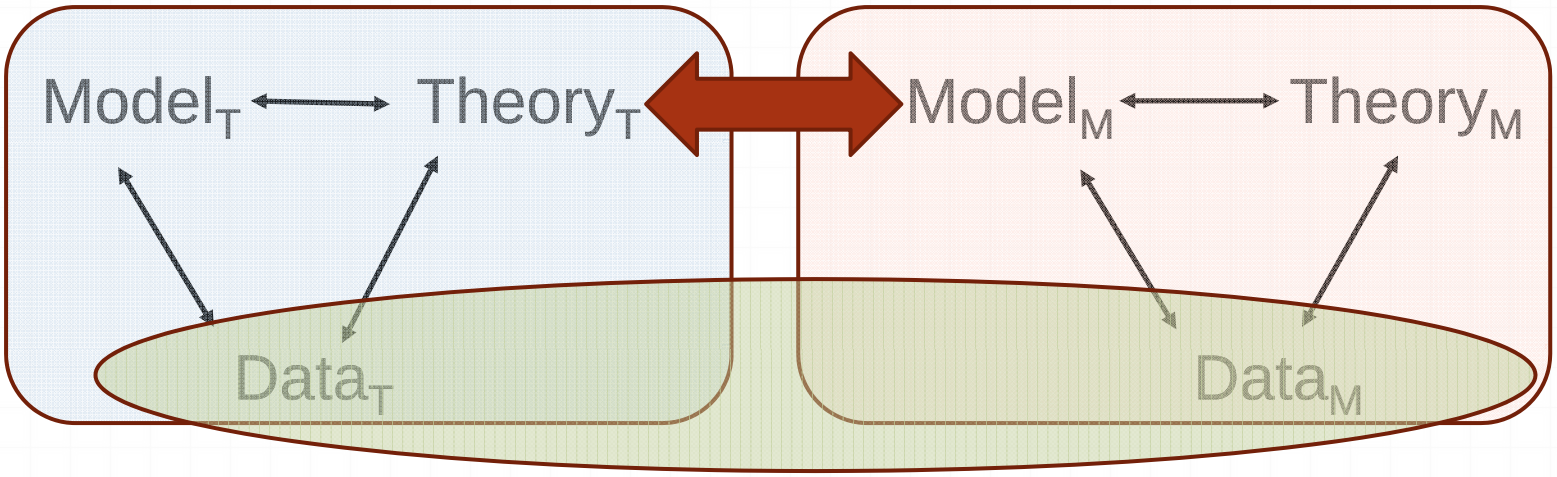
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Utility of the Theory-Model-Data Triangle

- Description of Interrelationships of T, M and D
- Normative Implications Based on Correspondences Among T, M, and D
  - Mapping of Meanings from One to the Other
  - Validity of Mappings
- Historical Implications
  - Which leads?
  - Interactions-Interchanges Among T, M and D
  - Changes in T, M and D as a Result of the Interactions

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

A More Accurate Picture (Probably)



Theory World

Data Modeling World

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Correspondences/Relationships

- Substantive Theory and Data Model ($T_T \longleftrightarrow M_M$)
- Substantive Theory and Theory of Data Modeling ($T_T \longleftrightarrow T_M$)
- Data in Theory World and Data in Model World ($D_T \longleftrightarrow D_M$)

Etc……

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Utility

- Systematic inquiry into and questioning of our assumptions about mappings
- Theories of mappings
- Awareness of historical nature of the research process—Unaware to us one vertex may be driving the research
- Theories of cycling among the vertices

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Q3. How Can Data Driven Research Serve Theory Driven Research?

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Moving From the Data World to the Theory World

- Typical Form of Theory: Causal Model (Flat)
- Alternative Types of Theory
  - Contingency Theory
  - Multilevel Theory
  - Process Theories

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Contingency Theory

- Concerned with effectiveness of an individual, group, organization, community, or society (any unit)
- Effectiveness is not caused or determined
- It depends…
- On the response of the organization to the demands of the situation (contingencies), which may vary from case to case

Contingency theory: Systematic framework for explaining effectiveness as a function of response to situations

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Format of Contingency Theory

1. Set of contingency variables
2. Set of possible responses to demands placed by contingencies
3. Set of outcomes
4. The "Contingency Equation":

   Positive outcomes will result if the response is appropriate for the contingencies (i.e. it "fits"). Negative outcomes will result if the response is not appropriate for the contingencies (i.e., it does not fit)

- To develop a contingency theory we must define which responses fit various combinations of contingency variables and so should be expected to yield positive outcomes
- The theory can be tested by assessing whether fit between contingencies and responses really is positively related to outcomes

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

To Help Contingency Theorists

- Mappings of contingencies and responses onto effectiveness

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Multi-Level Theory

- Assumes that there are causal factors at more than one level and that causes on one level cannot be reduced to causes on another
- Assumes causes at different levels may affect those on other levels
- Example: Educational effectiveness
  - Student (individual level)
  - Teacher (classroom level)
  - Curriculum (school level)
  - SES (community level)

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Format of Multi-Level Theory

- Lower levels are affected by higher level causes
- Individual student performance =
  f(individual characteristics, class, school, community)
- One approach: different regression slopes for subsets of classes, schools, and communities

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

To Help Multi-Level Theorists

- Observations of patterns and relationships on different levels
- Relationships across level

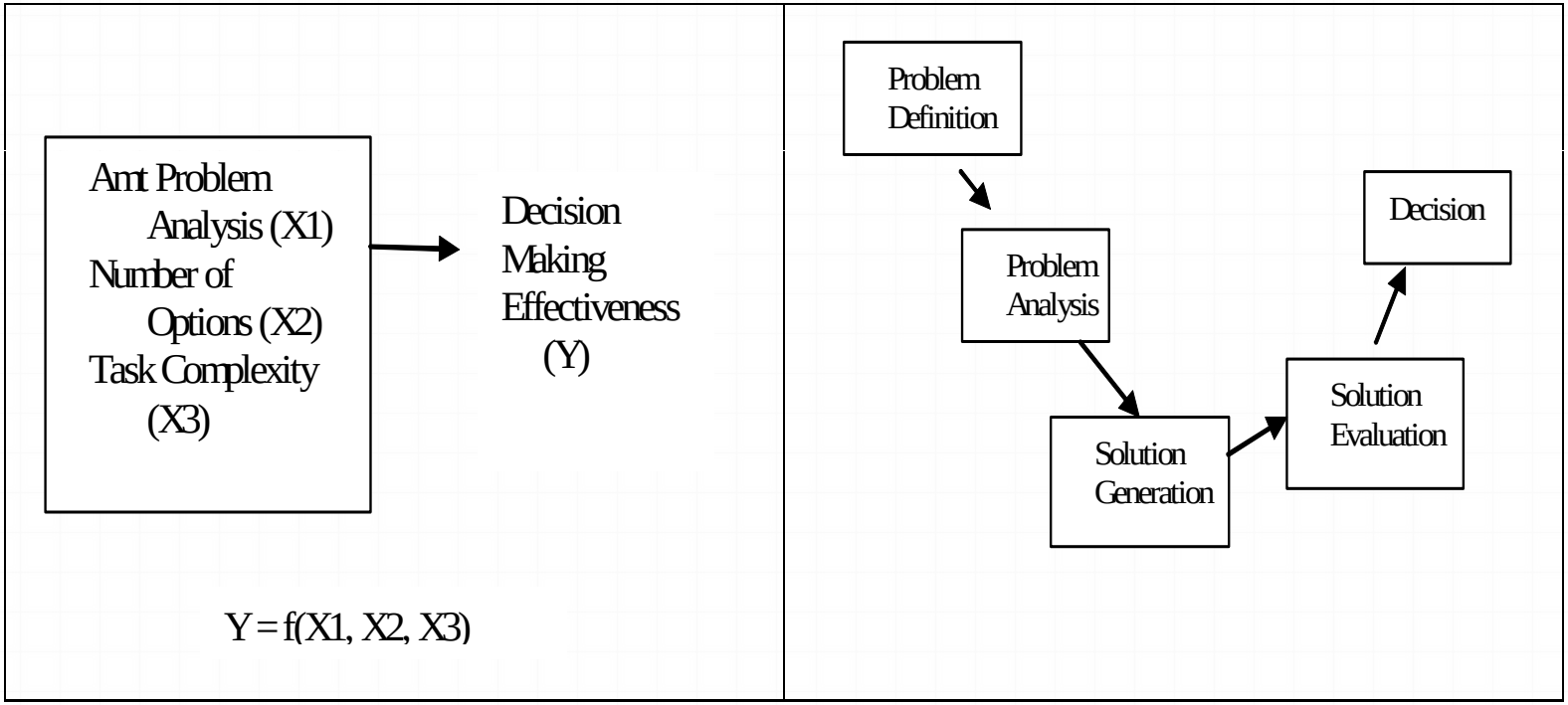
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Process Theory

- Focuses on unfolding of process over time
  - Life cycle
  - Dialectic
  - Evolutionary model (variation-selection-retention)
- Pattern identification
- Explanation via multiple elements: formal causation, critical events, historical junctures, etc.

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Process Theory



Variance Theory **Process Theory**

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

To Help Process Theorists

- Identify patterns over time
  - Stages are easy
  - Evolution and dialectics will be more difficult

- Critical events/Watershed events

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Q4. How Can Theory Driven Research Best Serve Data Driven Research?

Uhhh…..I dunno….

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

So Where Are We????

Abductive stage

Yet to be fully worked out….

Or contextualized to disciplines/areas….

Comments/Questions/Other Ideas?

Data-Driven Discovery

- Data Sources
  - Telescope/Large Instrument Analogs
  - Anthill Analogs
- Analytical Techniques
  - Transform/Clean/Arrange Data Into Useful Format
  - Pull Information From Data

Retroduction

Charles Sanders Peirce on Musement:

“There is a certain agreeable occupation of mind which…is not as commonly practiced as it deserves to be; for indulged in moderately -- say through some five to six per cent of one's waking time perhaps during a stroll -- it is refreshing enough more than to repay the expenditure. …In fact, it is Pure Play.

Now, Play, we all know, is a lively exercise of one's powers. Pure Play has no rules, except this very law of liberty. It bloweth where it listeth.

The particular occupation may take either the form of aesthetic contemplation, or that of distant castle-building (whether in Spain or within one's own moral training), or that of considering some wonder in one of the Universes, or some connection between two of the three, with speculation concerning its cause. It is this last kind -- I will call it "Musement"

'A Neglected Argument for the Reality of God', 1908