Lecture 6: Basics & Underlying Theory

Dewayne E Perry
ENS 623
perry@ece.utexas.edu

Discovery

- Process of Discovery
  - Plausible: interesting idea
  - Important: is it worthy of further consideration?
  - Acceptable: do we have a testable theory, can we create an hypothesis for experimental confrontation?
  - Justifiable: amenable to evaluation, defense, confirmation?
- Sources of Discovery
  - Intensive case studies
    - Document certain variables/conditions as prerequisite for a more theoretical study
  - Paradoxical incidents
    - Puzzled by contradictory aspects of a situation
  - Metaphors that stimulate our thinking
  - Rules of thumbs, folk wisdom
  - Account for conflicting results
    - Eg, performance in presence of others

Criteria

- Correspondence with reality
  - Idea agrees with what is true
  - Assumption: *speculative ideas that correspond with accepted truths have a higher payoff potential*
  - Use common sense, critical feedback and knowledge of the field
  - Problem: phantoms – accepted truths act as blinders
- Coherence and Parsimony
  - Coherence: whether it sticks together
  - Parsimony: spare and frugal
  - Occam's razor is a prescription for science
    - Eliminate what is not necessary
    - Similar to Einstein's principle: as simple as possible but no simpler

Criteria

- Falsifiability:
  - Must be refutable by observation; otherwise not scientific
  - Theory Y better than X if more falsifiable
  - BUT theories may evolve towards simplicity, not just falsifiability
  - Incomplete falsifiability may mean adjustment not discarding
- Agreement on falsification
  - Theory must be stated so that it can be falsified by a finite set of observations
  - A scientific theory can only be falsified, never proved correct
  - Theories may evolve as additions to as well as replacement of outmoded formulations
  - If an hypothesis does not receive support, it means the theory in its current form *may* be wrong
  - If a theory is repeatedly not supported, despite repeated attempts, it should be discarded
  - If an hypothesis *is* supported, the theory is not proved correct because another theory might account for all the existing results

Asking Questions

- Asking questions + systematic process to obtain valid answers
  - Make the question clear
  - Hypothesis should be consistent with questions
  - Statement of the problem
  - Critical: asking the right or important questions
- Types of Questions
  - Existence
  - Description/Classification
  - Composition
  - Relationships
  - Descriptive-Comparative
  - Causality
  - Causality-Comparative
  - Causality-Comparison Interactions

Types of Research Questions

- Existence questions
  - Does X exist? X is a thing, attribute, phenomenon, behavior, ability, condition, state of affairs etc.
    - Is there a tool that can generate X?
    - Is there a programmer who can write 200k lines per year?
  - Important when controversial
  - Generalization not important, existence is
  - Requires careful scientific work
  - Rule out alternative explanations
- Description/Classification
  - What is it like, is it variable or invariant, characteristic limits, unique of member of a known class, a distinctive description?
    - What are the limits of tool X?
    - What are the characteristics of structured programs?
  - Answer requires statements about:
    - Generality and representativeness of sample
    - Uniqueness/distinctness to population

Types of Research Questions

- **Composition**
  - What are the components of X?
    - What are the principle traits of a good programmer?
    - What are the main factors in a maintainable program?
  - Requires analysis or breakdown of whole into component parts
  - Factor analysis requires care and accuracy
  - Need large enough samples to rule out biases
- **Relationships**
  - What is the relationship between X and Y?
    - Are exceptions needed for maintainable programs?
    - Is elegance a function of age?
  - For predictiveness, can use multiple regression techniques
  - Or do the relationships fit theoretical models
  - Need valid/reliable measures, sufficient and representative samples, accurate computations, and interpretations supported by the data

Types of Research Questions

- Descriptive-Comparative
  - Is group X different from group Y?
    - Are Fortran programmers different from Lisp programmers?
    - Do novice C++ programmers make more errors than Java programmers? Experienced programmers?
  - An elaboration of the simple description question
  - Comparison may be organismic
    - Eg, age, weight, height
  - Comparison may be socio-economic
    - Eg, income, job, neighborhood
  - Must ensure equivalence of other characteristics
  - Criteria measures critical – need validity, reliability

Types of Research Questions

- Causality
  - Does X cause, lead to, or prevent changes in Y?
    - Does C++ lead to complex programs?
    - Does using exceptions lead to simpler programs?
  - Manipulate independent variables to get changes in dependent
  - Need control group for non-treatment
  - Must select sample carefully to rule out biases
  - Replications to warrant generality
- Causality-Comparative
  - Does X cause more change in Y than Z?
    - Is C++ better than Java in preventing race conditions?
    - Is the Jackson design method better than the Booch method in producing concurrent systems?
  - Compare rival treatments, control
  - Must guarantee that rival treatments are valid and are given in an unbiased manner

Types of Research Questions

- Causality-Comparison Interactions
  - Does X cause more changes in Y that Z under certain conditions but not others?
    - Do formal methods work better than informal methods for Europeans but not North Americans?
    - Is the MacOS easier to use than the Windows by naïve users but not experiences users?
  - Add more independent variables

Causation

- Philosophical issues
  - Aristotle: formal, material, efficient, final causes
  - Alternatives: concomitant variation, invariable sequence
  - Issue of power, causal efficacy -> invariably joined
  - Necessity versus invariable sequence
  - Sufficiency
  - Temporal precedence

Causation

- Plurality of causes
  - In behavioral sciences more in terms of one of the causes than a single cause
  - Experiment while holding everything else constant
  - Causality reserved for experimental results
  - Working definition of cause
    - Forget infinite regressive trail of reason
    - Practical (ie, non-philosophical) working definition:
      - ✓ *A proximal antecedent agent or agency that initiates a sequence of events that re necessary and sufficient in bringing about the observed effects*
    - Proximal since it is occurs at a time near the result
    - Antecedent as it clearly precedes the effect
    - Agent is set up intentionally
    - Experimenter exercises the control lever
    - Sufficient: effect not seen in absence of treatment

Lack of Causation

- Correlation and causation
  - Correlations show a relationship
  - With path analysis, multiple regression analysis, one can begin to make causal inferences, or build causal models
  - *But* demonstration of causality is a logical and experimental, rather than a statistical problem
- Enabling versus causing
  - Permits but does not cause
  - Eg: marriage is primary cause of divorce - NOT
- Other
  - Not sufficient
  - May not be necessary (eg, high IQ and good living)
  - May be probabilistic
- Some unsubstantiated causal claims:
  - Drinking wine prevents arteriosclerosis
  - Eating broccoli prevents colon cancer
  - *Post hoc ergo propter hoc* – necessity not demonstrated

Randomness

- Randomness is fundamental in experiments
  - Quasi-experimental otherwise
  - Must compensate for it otherwise
- Must have a clear view of its role
  - It is "the reasoned basis for inference" in experiments
  - The basis for statistical methods
- RA Fischer, *The Design of Experiments*, Edinburgh: Oliver & Boyd, 1935, 1949
  - Credited with the invention of randomization
  - Introduced the formal properties of randomization

The Lady Tasting Tea

- "A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup."
- Experiment
  - 8 cups of tea
  - 4 made each way
  - Presented in random order
    - often determined by a random number table
  - Subject knows the experimental design
  - Her task is to determine two sets of 4
    - Agreeing if possible with the treatments received

The Lady Tasting Tea

- What would be expected if the Lady was "without any faculty of discrimination"
  - Ie, if she made no changes in her judgments in response to changes in the order of presentation
  - There are 70 possible divisions of 8 into 4
- Randomization has insured that all orderings are equally probably
- The chance of accidentally choosing the right ordering is 1/70
  - Ie, probability of random ordering agreeing with the Lady's fixed judgments is 1/70 or $\binom{8}{4} = 70$
  - 0.014 is the significance level for testing the null hypothesis of having no judgment

The Lady Tasting Tea

- Example serves well
  - The Lady is not a sample from a population of ladies – concerns her alone
  - Her eight judgments are not independent observations (rule of 4 each)
  - Later cups differ from earlier ones
- Inferences are justified because the only probability distribution used in the inference is the one created by the experimenter

Fischer's argument

- Experiments do not require
  - That experimental units be homogeneous
  - That experimental units be a random sample from a population of unit
- It is sufficient to require that treatments be allocated at random to experimental units for valid inferences
- Probability enters the experiment only thru the random assignment of treatments

The "No Effect" Hypotheses

- Special place to the test of the hypothesis that the treatment is entirely without effect
- Reason: in a randomized experiment, this test may be performed virtually without assumptions of any kind – ie, relying only on random assignment
- Contribution of randomization is clearest when expressed in terms of the test of no effect
  - Does not mean that such tests are of greater practical importance
  - It sets randomized and non-randomized aspects in sharper contrast
- Whereas inferences in non-randomized experiments require assumptions

The “No Effect” Hypotheses

- To say a treatment has no effect is to say that that each unit would exhibit the same value of the response whether assigned to treatment or to control
- A change in response indicates the treatment has some effect
- Will discuss later various tests of the significance and the size of effects

Alternate Hypotheses

- As opposed to the null hypothesis, experimental hypotheses take a stand
  - Different treatments behave differently (two tailed prediction)
  - Predict what direction the expected differences take (one tailed)

Hypotheses and Theory

- Theory is a large scale map, different areas represent general principles
- Hypotheses are like small sectional maps, focus on specific areas
- Conceptual similarities
  - Range from very explicit to very vague
  - Fall back on hidden assumptions, regulative principles
  - Give directions to our observations
- Some hypotheses spring from experimental observations
  - Don't know where they will go
- Others from theory
  - Conceptual hypotheses
  - Rely on previous studies, theories

Creating Hypotheses

- Defining terms
  - For observations to have value, abstractions have to be concretized
  - Two types of definitions
    - Operational: x is defined in terms of test y under conditions z
    - Theoretical: abstract constructs used
  - At some point must be operational for the experiment
  - Hypotheses are predictive statements about the expected outcome
  - They call for a test and embed a conclusion
  - Explicit statements are *de rigueur*
  - When comparisons are predicted they have to be explicated

Stating the Hypothesis

- Concomitant variation: X is a direct function of Y
- Comparative: other things being equal ….
- H0 and H1 (null and alternative) are mutually exclusive and exhaustive
  - Usually a specific H0 and a general H1
  - Try to reject H0

Hypothesis Testing Errors

- 2 types of errors: I and II
  - Type I – rejecting the null hypothesis when it is true
    - Greater psychologically important risk
    - Think there is a relationship when there is not
    - Waste time in blind alleys
  - Type II – accepting the null hypothesis when it is false
    - Deny a relationship when there is one
    - In effect, reject useful results