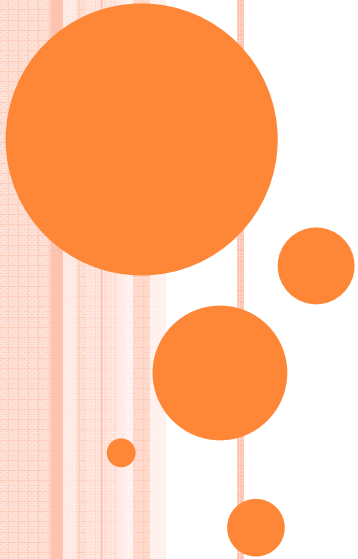


QUANTITATIVE EVALUATION

Concerns of Designing Controlled Experiments

SCIENTIFIC APPROACH

- Beyond “user friendly”
- Specify users and tasks
- Predict and measure
  - time to learn
  - speed of performance
  - rate of human errors
  - human retention over time
  - subjective satisfaction
- Accommodate individual differences
- Consider social, organizational & cultural context

QUANTITATIVE EVALUATION

- Collect (performance) measurements

- Methods
  - User events collection
    - # of mouse clicks, # of keys depressed, mouse moving distance
    - data collected during system use
      - Google Analytics
  - Controlled experiments
    - lucid/testable hypothesis
    - independent variables
    - dependent variables
    - can be reproduced by others

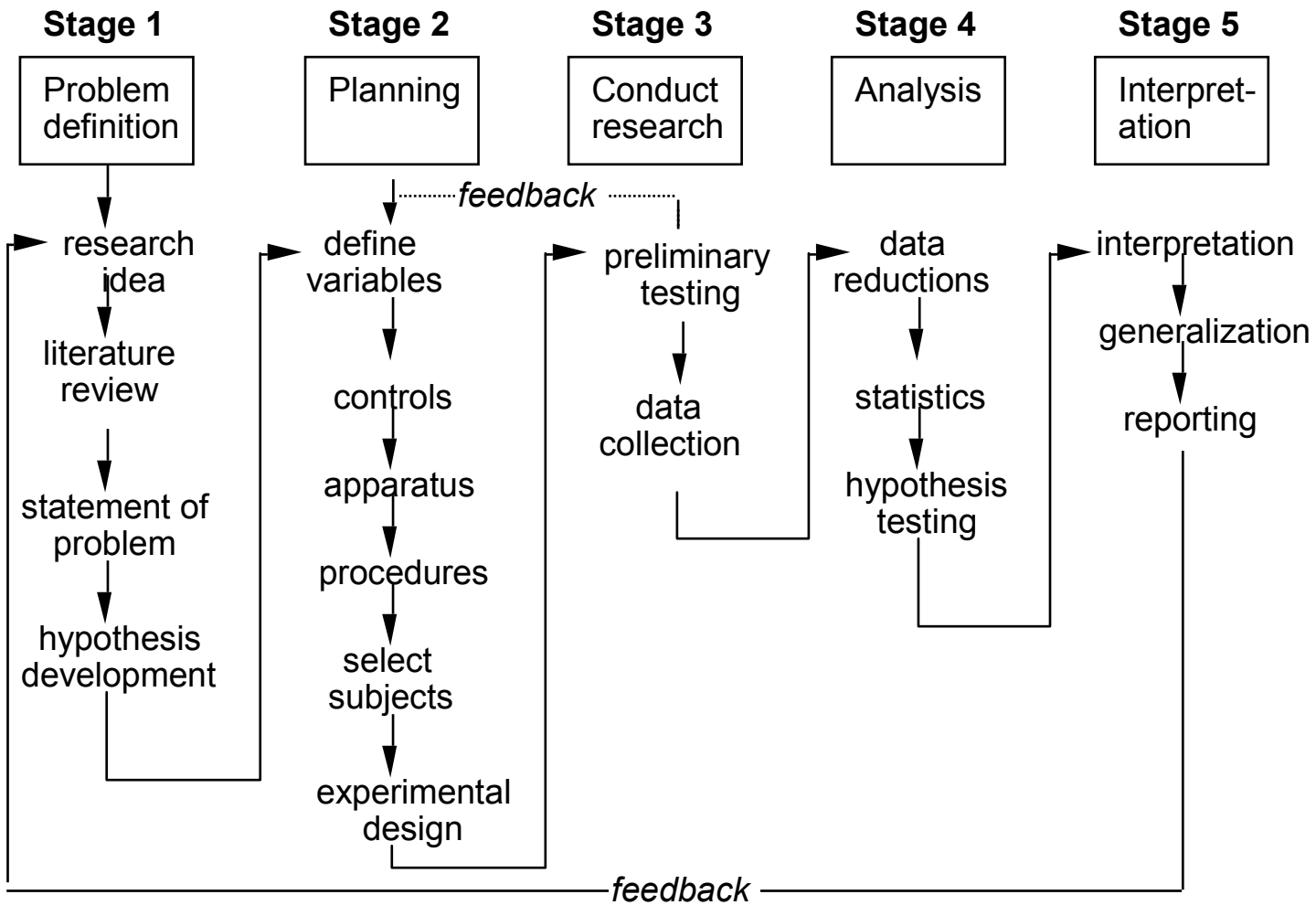
CONTROLLED EXPERIMENTS

- Based on Practical problem & Existing theory
- Outcome
  - guidance for practitioners
  - refine theory
  - advice for experimenters
- Carefully select and assign subjects
- Control other variables

CONTROLLED EXPERIMENTS

- Start with a lucid and testable ***hypothesis***
  - "interaction method A is faster than method B"
- Alter small number of ***independent variables***
  - interaction methods, predefined user classes, tasks
- Measure small number of ***dependent variables***
  - speed in time, errors, subjective satisfaction
- Apply ***statistical hypothesis test***
  - t-Test, Anova

PLANNING FLOWCHART FOR EXPERIMENTS



from an early ACM CHI tutorial

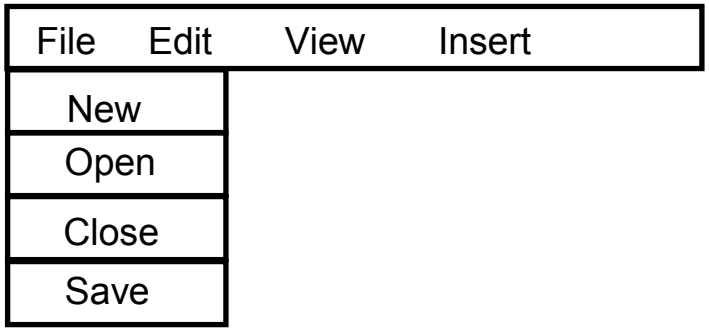
NOTES FOR RUNNING THE EXPERIMENT

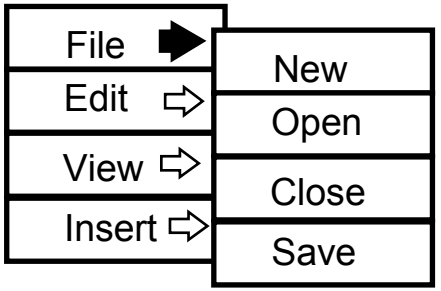
- Always run ***pilots*** first!
  - There are always unexpected problem!
  - When the experiment has started you cannot change
- Use a ***check-list*** so that all subjects follow the same steps
- Don't forget the ***consent form***!
- Don't forget to ***debrief*** each subject

LUCID AND TESTABLE HYPOTHESIS

- Example:

  There is no difference in user performance (time and error rate) when selecting a single item from a pop-up or a pull down menu of 4 items, regardless of the subject's previous expertise in using a mouse or using the different menu types"





Adapted from Saul Greenberg's slide

INDEPENDENT VARIABLES

- the things you manipulate independent of a subject's behavior
- determines a modification to the conditions the subjects undergo
- may arise from subjects being classified into different groups
- example:
  - menu type: pop-up or pull-down
  - menu length: 3, 6, 9, 12, 15
  - subject type (expert or novice)

Adapted from Saul Greenberg's slide

DEPENDANT VARIABLES

- variables dependent on the subject's behavior or reaction to the independent variable
- the specific things you set out to quantitatively measure
- example:
  - time to select an item
  - selection errors made
  - time to learn to use it to proficiency

Adapted from Saul Greenberg's slide

SUBJECT SELECTION

- judiciously select and assign subjects to groups
- ways of controlling subject variability
  - reasonable amount of subjects
  - random assignment
  - make different user groups an independent variable
  - screen for anomalies in subject group
    - superstars versus poor performers

Adapted from Saul Greenberg's slide

CONTROLLING BIAS

- unbiased instructions
- unbiased experimental protocols
  - prepare scripts ahead of time
- unbiased subject selection

Adapted from Saul Greenberg's slide

CONFOUNDING VARIABLES

- correlates with both dependent & independent variables
- cause Type I errors: false positive
- hurts internal validity
  - observed effect is not attributed to independent variable

- example
  - GDP rise and child's weight

- how to control
  - case-control study, cohort study, stratification
  - randomization on sufficiently large number of subjects

BLOCKING

- arrange experimental units in groups (blocks) that are similar to one another within group than the whole set
- reduce variability → greater precision
- e.g., 10 volunteers and two different types of shoe soles
  - randomly assign 5 and 5
  - left foot and right foot : each person is a block
- randomize within the block
- measure difference within the block

BETWEEN SUBJECTS DESIGN

- Distinctive subjects group for each treatment
- Groups are not blocks
  - random assignment
  - only one treatment for each group
- Observed results are compared between groups
- Can eliminate ordering/learning effects

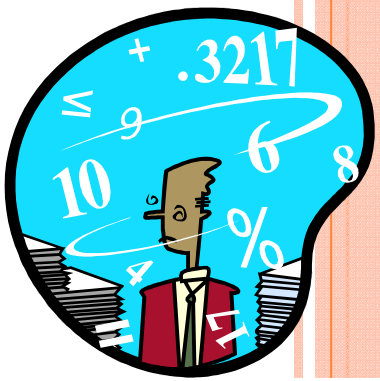
- Might invite confounders
- Need more subjects
- Difference between subjects might introduce a bias

WITHIN SUBJECTS DESIGN

- Same subjects group for all treatments
- Observed results are compared within each subject
  - pared t-Test
- Can eliminate variation due to subject differences
- Might suffer from ordering/learning effects
  - counter balance by randomizing the treatment order
- Need fewer subjects

STATISTICAL ANALYSIS

- Apply statistical methods to data analysis
  - confidence limits:
    - the confidence that your conclusion is correct
    - level of statistical significance : 0.05 or 0.01
    - test of significance → p-value
    - "the hypothesis that technique X is faster is accepted" means:
      - a 95% chance that the hypothesis is true
      - a 5% chance you are wrong



Adapted from Saul Greenberg's slide

IRB: Institutional Review Board

- IEC: independent ethics committee
- ERB: ethical review board

- A committee to approve, monitor, and review research involving human subjects with the aim to protect the rights and welfare of the research subjects, *Wikipedia*

- biomedical and behavioral research

- Research Act of 1974, USA

When is IRB Review Required?

- IRB review is required for research involving human subjects.

- What is a **human subject**?
  - A living individual about whom an investigator (whether professional or student) conducting research obtains
    - data through intervention or interaction with the individual, or
    - identifiable private information
- What is **research**?
  - "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge."

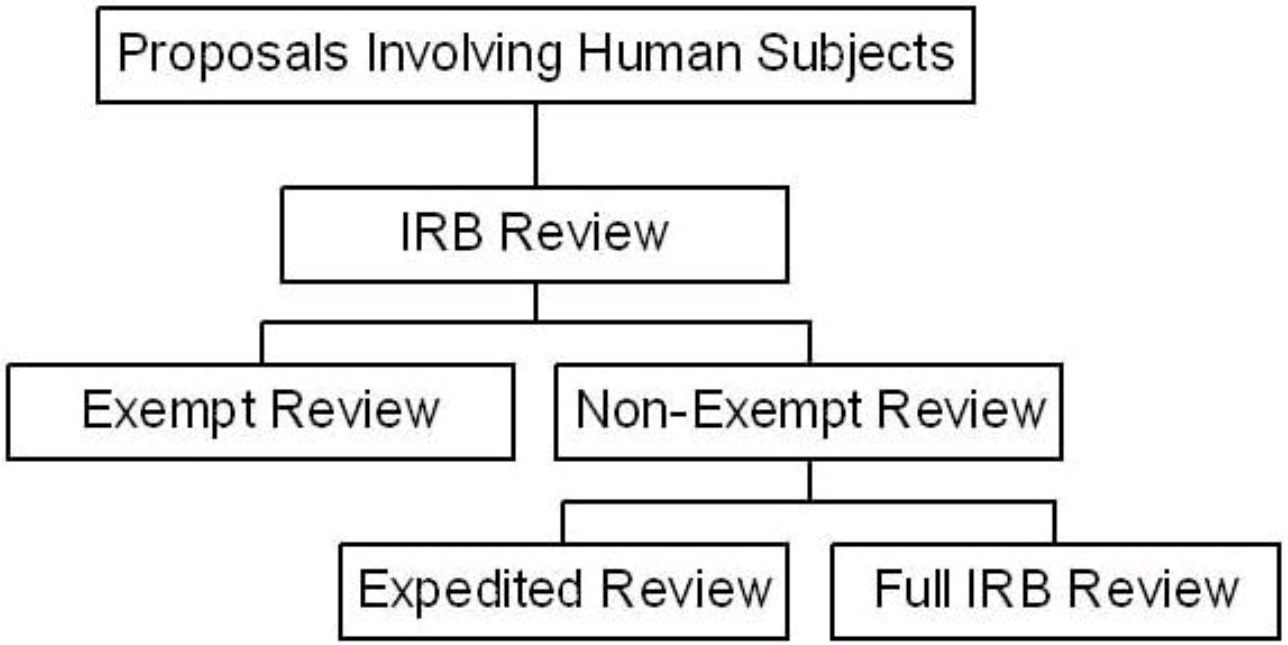
서울대 생명윤리 심의 위원회

- 생명윤리 및 안전에 관한 법률 시행, 2005.1.1

- 인간 대상의 연구에서 피험자 등 연구에 참여한 사람들의 존엄성, 권리, 안전, 복지를 증진시키고 과학적 연구의 신뢰를 높이기 위함
- 대학을 위시한 연구기관 내에 배아연구, 유전자연구 등 생명과학 관련 특정연구의 윤리성을 심의하는 기관생명윤리심의위원회를 설치하여 운영할 것이 의무화
- 현재 요청 시 심의를 실시함 - 곧 의무화 될 예정

- http://irb.snu.ac.kr/

IRB Review Process



http://www.umbi.umd.edu/research- development/compliance/images/irb- diagram.jpg

INFORMED CONSENT FORM

Project Title	*Evaluation of an Interactive Exploratory Analysis Tool (The Hierarchical Clustering Explorer)*
Statement of Age of Subject	*I state that I am over 18 years of age, in good physical health, and wish to participate in a program of research being conducted by Prof. Ben Shneiderman in the Department of Computer Science at the University of Maryland, College Park.*
Purpose	*The purpose of this research is to evaluate the hierarchical clustering explorer and to improve it though iterative design process.*
Procedures	*This experiment consists of three parts; using the tool for a week, meeting with the investigator, and answering a questionnaire. I will be asked to use the software at least 30 minutes a week during a regular meeting time with the investigator. During the meeting, I'll be asked to explain what I did with the tool for a week, and I will show the investigator how I use the tool to analyze my data. Questions that the investigator will ask during the weekly meeting may include the following:* *- What kind of ranking criteria are most frequently used?* *- What are the missing criteria that users most want?* *- Does the score overview help users identify interesting projections?* *- Does the histogram/scatterplot browser help users traverse projections?* *- Do the search mechanisms in the parallel coordinates view help users identify interesting items?* *- What are the most frequently used search mechanisms in the parallel coordinates view?* *I'll also talk to him about possible improvement, if any, that I identified using the tool. I may decline to answer any of the questions and I will not be penalized in any way.*
Confidentiality	*All information collected in this study is confidential to the extent permitted by law. I understand that the data I provide will be grouped with data others provide for reporting and presentation and that my name will not be used.*
Risks	*Participation involves risks that are no greater than those encountered in ordinary daily living. There is no risk except for using specific software (the hierarchical clustering explorer) to analyze my data set at least 30 minutes a week during the experiment period.*
Benefits, Freedom to Withdraw, & Ability to Ask Questions	*The experiment is not designed to help me personally, but to help the investigator learn more about the strength and weakness of the hierarchical clustering explorer in a realistic environment. I am free to ask questions or withdraw from participation at any time and without penalty.*
Contact Information Of Investigators	*Prof. Ben Shneiderman* *3177 A.V. Williams Bldg. Department of Computer Science* *University of Maryland, College Park, MD 20742* *Phone: 301-405-2680* *Email: ben@cs.umd.edu*
Please add name, signature, and date lines to the final page of your consent form	NAME OF SUBJECT _____ SIGNATURE OF SUBJECT _____ DATE _____

Contact Information of Institutional Review Board:
If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742;
(e-mail) **irb@deans.umd.edu**; (telephone) **301-405-4212.**

What we know now

- Quantitative evaluation methods
  - User events collection
  - Controlled experiments
- Controlled experiments
  - lucid and testable ***hypothesis***
  - Alter small number of ***independent variables***
  - Measure small number of ***dependent variables***
  - Apply ***statistical hypothesis test***
- Experimental Design
  - Within/between subjects design, blocking, confounding variables
- IRB