

# Evaluation

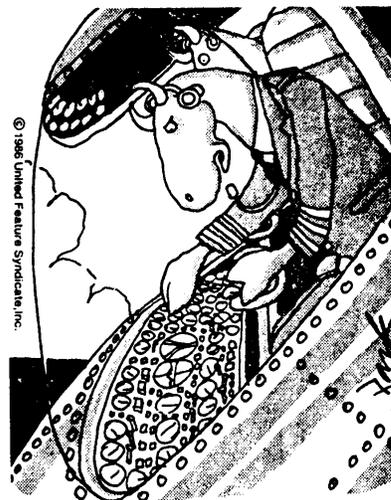
CMSC 436/636

Penny Rheingans

University of Maryland, Baltimore County

## Evaluation of Visualization

- Is this technique useful?
  - For what?
  - To whom?
- Why is this technique useful?
- What other techniques might be useful?
- How does visualization change the nature of thinking / learning / decision-making?



"Darn these hooves! I hit the wrong switch again! Who designs these instrument panels, raccoons?"



## Performance Evaluation

- What are system requirements?
- How long will it take to run?
- How big a problem can I solve?



## Quantitative User Studies

- Measure metrics on performance of an abstract task
  - Accuracy
  - Time
  - Confidence

## Experiment

- **Experiment**
  - Def: procedure in which active independent variable is manipulated and changes in the dependent variable are measured under controlled conditions
  - Experimental operational definition: specifies operations used to manipulate a conceptual independent variable
  - Measured operational definition: specifies procedures used to measure a conceptual variable
- **Hypothesis**
  - Def: statement of tentative relationship between antecedent condition and subsequent behavior
  - Characteristics:
    - must be an explicit statement of the relationship
    - must be empirically testable
    - must be formulated prior to empirical testing
  - Null hypothesis : statement that there is no relationship between condition and behavior

## Variability

- **Maximize effects of experimental variables**
  - use extreme values
  - use more than one value
- **Minimize effects of extraneous variation**
  - Sources of extraneous variation
    - attributes of subjects
    - procedural variations
    - physical environment
    - learning
    - trial order
  - Confounding: extraneous variable is systematically manipulated so that its levels covary with the independent variable
  - Strategies for minimizing extraneous variability
    - Hold extraneous variables constant
    - Balance effects of extraneous variables
    - Match participants on an extraneous attribute variable
    - Randomize extraneous variables
    - Incorporate extraneous variables as independent variables

## Experiment Design

- **Between-subjects Design**
  - Def: different groups of individuals receive different treatments; all subjects exposed to just one level of variable
  - Minimizing effects of subject differences
    - randomize assignment of subjects to treatments
    - match subjects on extraneous variable
- **Within-subjects Design**
  - Def: one group of subjects is exposed to all levels of each independent variable; all subjects receive all levels
  - Advantages
    - more data per subject
    - possibly greater sensitivity
    - can study interactions
  - Minimizing multiple treatment effects
    - hold conditions constant
    - balance conditions

## Statistical Hypothesis Testing

- **Basic procedure**
  - compute test statistic
  - accept or reject null hypothesis
- **p-value**
  - probability of obtaining a test statistic as extreme or more extreme than that observed, given that the null hypothesis is true
  - .05 standard for statistically significant difference
- **Possible errors**
  - Type I: reject null hypothesis when it's true (false positive)
  - Type II: fail to reject false null hypothesis (false negative)

## Analysis of Variance

- Partitioning observed variation according to source
  - obs value = constant effects + sum of factor effects + sum of extraneous effects
- Assumptions
  - normality of distribution of dependent variable values
  - equal variances
  - independence of scores
- Between-subjects ANOVA
  - ind score = grand mean + deviation of group mean + deviation from group mean
  - $F : MS_{\text{between}} / MS_{\text{within}}$
- Within-subjects ANOVA
  - one-factor score = grand mean + effect of treatment A + individual differences + effects of treatments by subjects interaction
  - two-factor score = grand mean + effect A + effect B + individual differences +  $A \times B + A \times S + B \times S + A \times B \times S$
  - $F_A : MSA / MSA \times S$

## Experimental Validity

- Statistical conclusion validity
  - were correct decisions made about null and alternate hypothesis?
- Internal validity
  - is observed effect due to manipulations done to independent variable?
- Construct validity
  - do measured qualities really indicate conceptual variable?
- External validity
  - can results be generalized to larger population?



## Qualitative Evaluation

- Anecdotal evidence from satisfied users
- Size of user community
- Qualitative user studies
  - Ethnographic analysis of work practices
  - Longitudinal field studies
  - Usability studies
- Analysis of design choices in conceptual framework