CS3205 – HCI in Software Development

More on Evaluation

More from Chapter 13 in published book

Tom Horton

*** Material from: [Floryan, Uva; Klemmer, UCSD]**

Where User Evaluation can Happen

- Controlled settings involving users
  - e.g. usability testing & experiments in laboratories and living labs.
- Natural settings involving users
  - e.g. field studies and in the wild studies to see how the product is used in the real world.
- Settings not involving users
  - e.g. to predict, analyze & model aspects of the interface analytics.

What You Can Do For Each Type of Evaluation

Method	Controlled settings	Natural settings	Without users
Observing	x	x	
Asking users	x	x	
Asking experts		x	x
Testing	x		
Modeling			x

Participant Rights and Informed Consent

- Participants need to be told why the evaluation is being done, what they will be asked to do and their rights.
- Informed consent forms provide this information.
- The design of the informed consent form, the evaluation process, data analysis and data storage methods are typically approved by a high authority, e.g. an Institutional Review Board.

Informed Consent

- Let's look at two examples of informed consent forms
  - Example of student study for usability work, approved by UVa IRB
  - Example from U of Ottawa researcher

Starting An Evaluation

- MUST have a research question!
  - Typically, these research questions are related to your usability requirements!
  - 'Which design accomplishes usability requirement X best!"
- It is very common to ask things like "Do you like my interface?"
  - Anything wrong with just asking people this?

Is Interface X better than interface Y?

- What does “better” mean?
  - Need explicit measures when possible.
- Nearly always, your answer will be that ‘It Depends’.
  - So, more interestingly, what does it depend on?
  - Figure that out! And then describe your results!

A FEW TERMS:

- Independent Variables
  - "Variable" that is being manipulated to study an effect via a change
- Dependent Variables
  - Variable that is measured for change after Independent Variable is altered.
- An example:
  - Changing interface from entering commands to voice recognition
  - Time to complete a task
  - Number of errors
  - Satisfaction

Controlled Comparison Enables Causal Inference

- When possible, control as many external variables as possible.
- Possible ways of doing this might include:
  - Users placed in groups randomly.
  - Users perform experiment in exact same environment.
  - Users are given the exact same instructions, training time, etc.
  - Any outside variable that could effect the result of the study should be equivalent for ALL test subjects.

Avoid Pitfalls in Evaluating Results

- Possible problems related to experimental design:
  - Reliability: does the method produce the same results on separate occasions?
  - Validity: does the method measure what it is intended to measure?
  - Ecological validity: does the environment of the evaluation distort the results?
    - Example: the Hawthorne Effect (the "observer effect")
  - Biases: Are there biases that distort the results?
  - Scope: How generalizable are the results?

Internal Validity

- Can you reproduce the experiment multiple times yourself!?
  - Same prototypes
  - Different users
  - Same experimental setup, conditions, etc.

External Validity

- Does your experiment apply generally to other 'outside' settings?
  - Different users selected from a different "pool".
  - Different prototypes with same general independent and dependent variables.
  - Different designers running the experiments.
  - Etc…
- In short, External Validity means your results apply generally to experiments with the same abstract characteristics as yours.

Bias and Leading Questions

- That question 'Do you like my interface?' is a leading question!
- A ***Leading Question*** is a question that suggests the answer the examiner is looking for or contains the information the examiner is looking to have confirmed.
- Don't ask users leading questions!

Please the Experimenter Bias

- People want to make you feel good about your work (because it's assumed that you worked hard).
- So users will tend to say 'Yes' to this question.
  - How do we get around this?

Getting Beyond "Do you like my interface?"

- Ways to get around "please the experimenter" bias
  - 1) Double-blind studies
    - i.e., Both user and facilitator don't know which experimental group the user is in.
  - 2) Don't let the user know what you are measuring / what you care about (until study is over).
    - E.g., Don't tell the user that the number of mistakes while typing is being measured.
  - 3) Ask questions that cancel each other out
    - E.g., Ask about how useful the interface was AND how frustrating it is. User can't tell which you care about.

Getting Beyond "Do you like my interface?"

- If possible, evaluation measures (quantifiable variables) should ALWAYS have a base-rate.

- Base-rates: How often does 'Y' occur in the current setting (if one exists)?
  - Very reasonable for some of your projects, if there is a competing or existing product that exists.

Getting Beyond "Do you like my interface?"

- Base-rates: How often does 'Y' occur in the current setting (if one exists)?

- Example: "User will make less than 3 mistakes while performing task X"
  - Where did the 3 come from?
  - If 3 is the average that users make on some competing system, than that is a good base-rate!
  - If 3 was made up, then it is not. It provides little context.
  - Note: We need to define what 'mistake' means clearly…but that's another story.

Getting Beyond "Do you like my interface?"

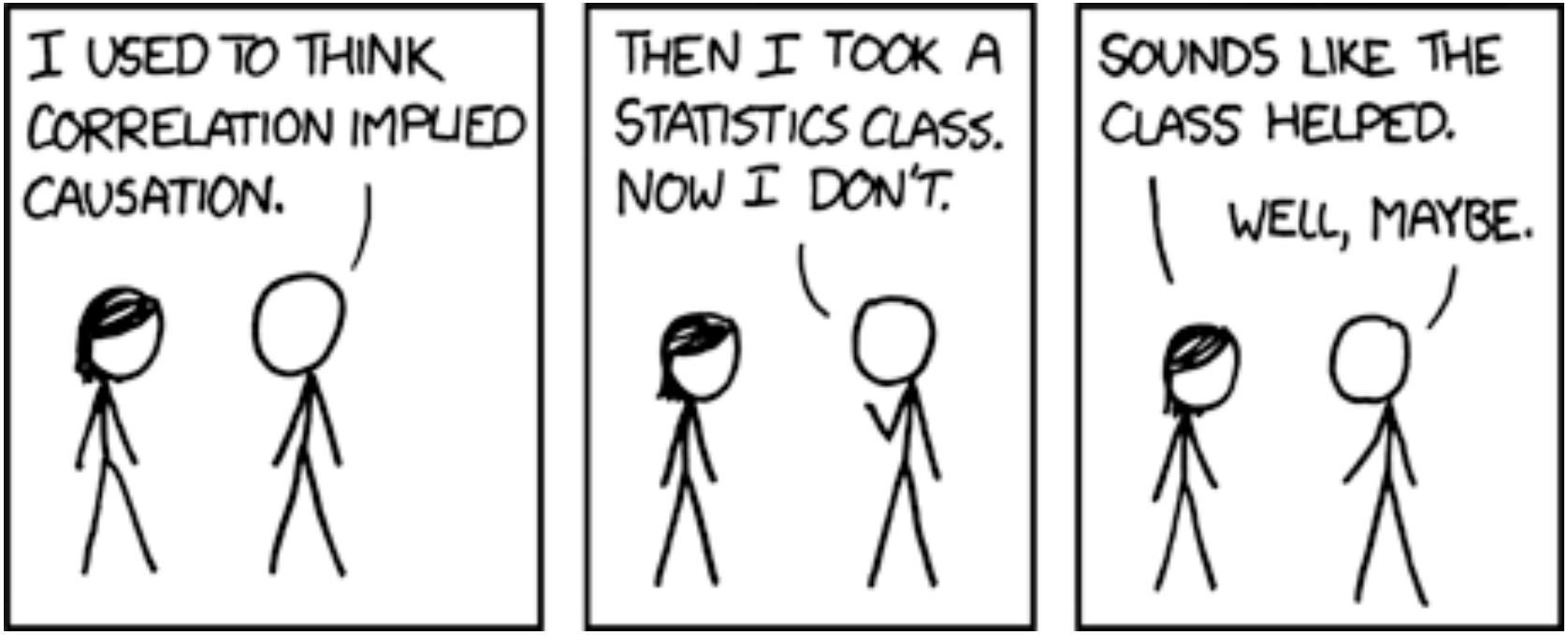
- Correlations: Do X and Y co-vary?
  - Requires measuring X and Y.
  - Probably need two prototypes OR two versions of a prototype (each with different X).
  - Note: here X is probably our independent variable, and Y is our dependent variable.

- Causes: Does X cause Y?
  - Requires measuring X and Y (establishing correlation).
  - Requires establishing time precedence.
  - Requires controlling for all confounding variables.

Examples of Correlation != Causation

- http://www.buzzfeed.com/kjh2110/the-10-most-bizarre-correlations

Correlation != Causation



*from http://xkcd.com/552/

Case Studies

- Two in textbook:
  - Computer Game
  - Skiers
- One from a paper
  - Children's Digital Library paper: "Supporting Elementary-Age Children's Searching and Browsing…"
    http://hcil2.cs.umd.edu/trs/2008-31/2008-31.pdf

The language of evaluation

- Analytics
- Analytical evaluation
- Biases
- Controlled experiment
- Crowdsourcing
- Ecological validity
- Expert review or crit
- Field study
- Formative evaluation
- Heuristic evaluation
- Informed consent form
- In the wild evaluation
- Living laboratory
- Predictive evaluation
- Reliability
- Scope
- Summative evaluation
- Usability laboratory
- User studies
- Usability testing
- Users or participants
- Validity

Questions?