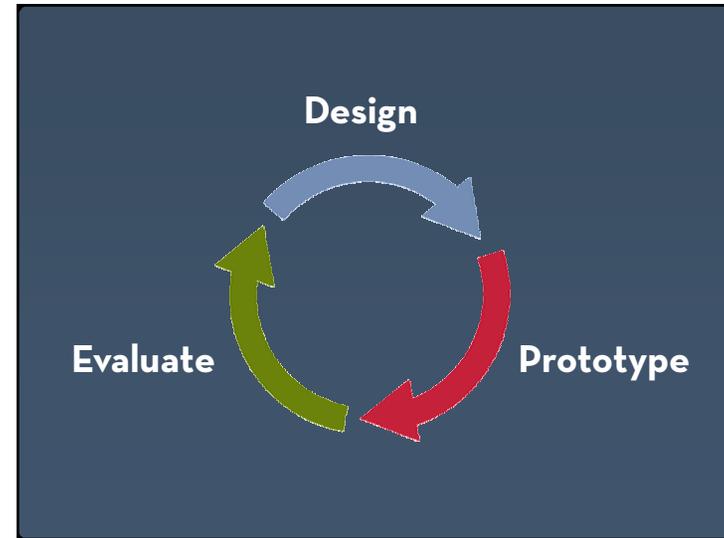




Human-Computer Interaction Design Studio

13 February 2012

<http://cs247.stanford.edu>



What is Evaluation?

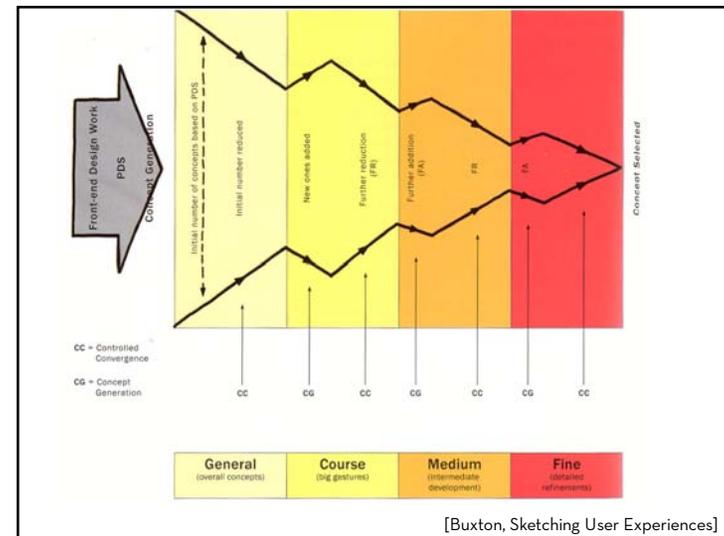
Part of the design-prototype-evaluate cycle.

Formative: Something you do during the design of a system to guide you on the right path.

Summative: Something you do at the end of a project to show that your design works.

Posing a question to the world so as to get a reliable and/or insightful answer.

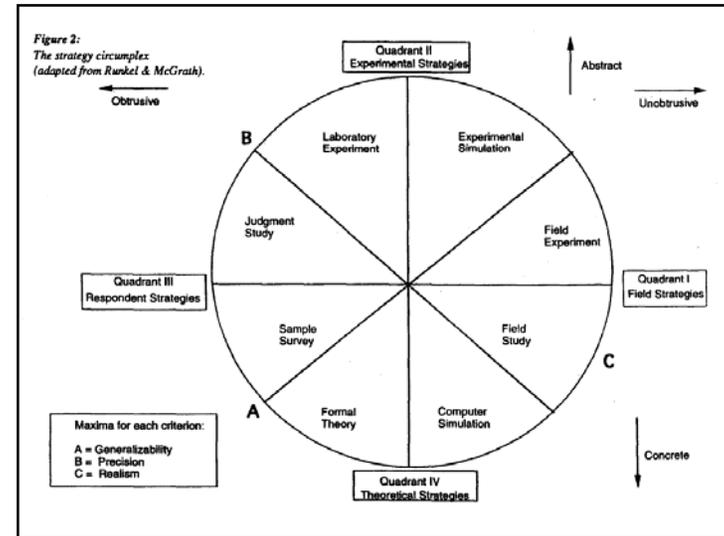
Why? So you can improve? Publish? Sell?



“Methods for establishing validity vary depending on the nature of the contribution. They may involve empirical work in the laboratory or the field, the description of rationales for design decisions and approaches, applications of analytical techniques, or ‘proof of concept’ system implementations.”

- ACM CHI Website

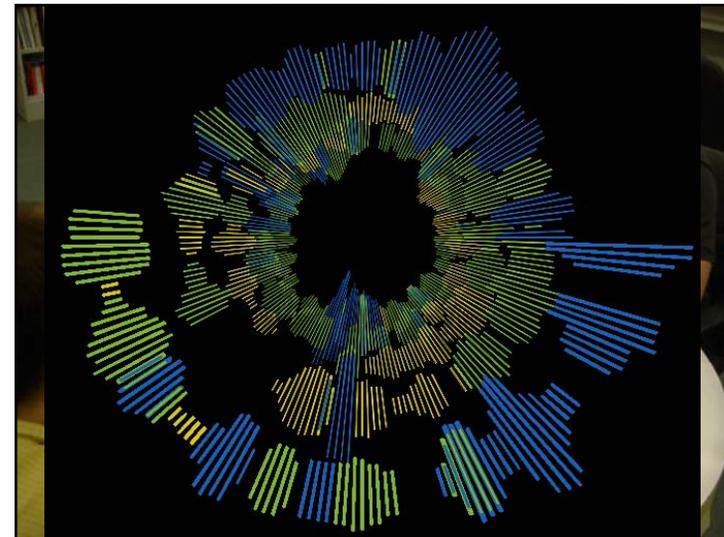
5

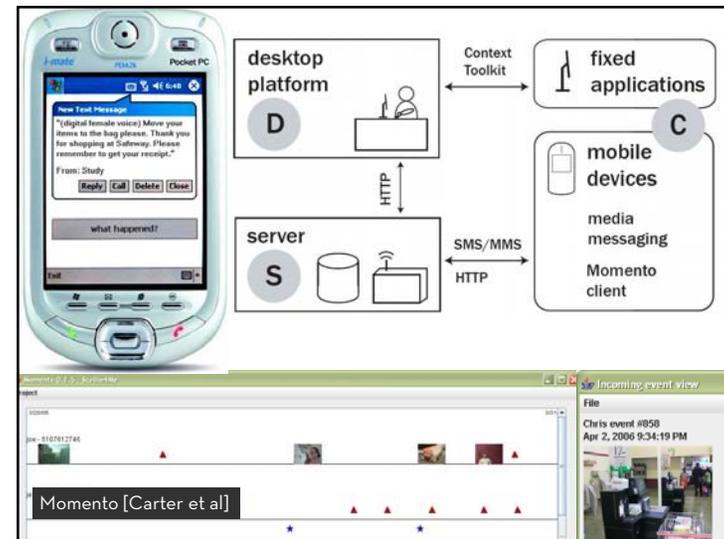
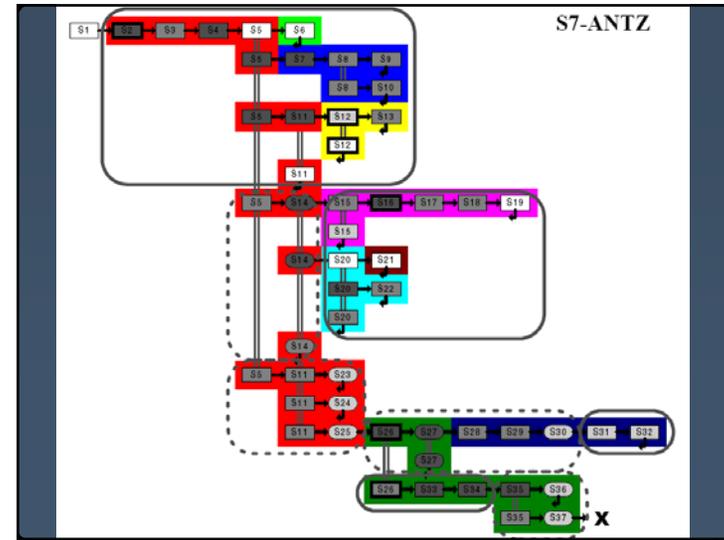
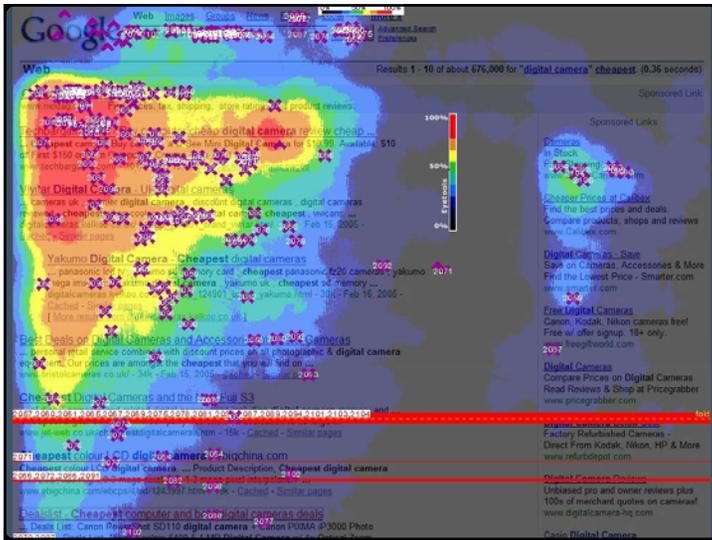


What to evaluate?

- Enable previously difficult/impossible tasks
- Improve task performance or outcome
- Modify/influence behavior
- Improve ease-of-use, user satisfaction
- User experience
- Sell more widgets

What is the motivating goal?





Evaluation Methods in HCI

Inspection (Walkthrough) Methods
Observation, User Studies
Experience Sampling
Interviews and Surveys
Usage Logging
Controlled Experimentation
Fieldwork, Ethnography
Mixed-Methods Approaches

Proof by Demonstration

Prove feasibility by building prototype system
Demonstrate that the system enables task
Small user study may add little insight



“You Are Not the User”

Seems obvious, but you may have
Different experiences
Different terminology
Different ways of looking at the world

Easy to think of self as typical user.
Easy to make mistaken assumptions!

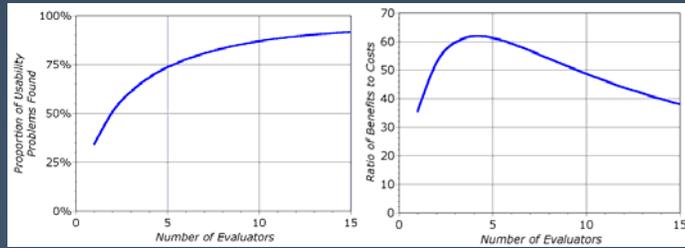
Inspection Methods

Often called “discount usability techniques”
“Expert review” of user interface design

Heuristic Evaluation (Nielsen, useit.com/papers/heuristic)

Visibility of system status	Recognition over recall
Match between system and real world	Flexibility and efficiency of use
User control and freedom	Aesthetic and minimalist design
Consistency and standards	Help users recognize, diagnose, and recover from errors
Error prevention	Help and documentation

How many evaluators?



17

Usability Testing

Observe people interacting with prototype

May include:

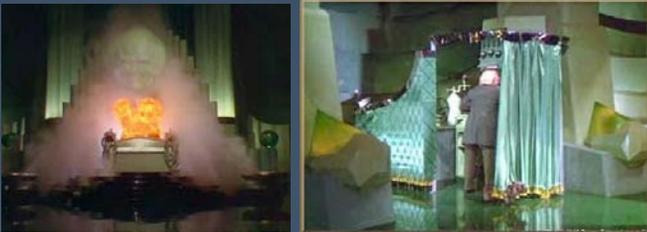
- Providing tasks (e.g., easy, medium, hard)
- Talk-aloud protocol (users verbal reports)
- Usage logging
- Pre/post study surveys
 - NASA TLX: workload assessment survey
 - QUIS: user interaction satisfaction

Wizard-of-Oz User Studies



Wizard-of-Oz Prototyping

To prototype an interactive system by using human operators to simulate machine behavior.



How to make a WOz Prototype

Map out scenarios and application flow: what should happen in response to user behavior?

Build interface “skeletons” (minimal autonomy)

Develop “hooks” for wizard input

Where and how the wizard will provide input (e.g., selecting next screen, entering text, recognizing speech or gestures); Must be possible to replace later with computer

Rehearse wizard role with team mates. The Wizard should be able to perform the task.

Tradeoffs in Wizard of Oz

Advantages

- Fast (faster), cheaper and more iterative prototypes
- Create multiple variations
- Identifies bugs and problems with current design
- Places the user at the center of development
- Can envision challenging-to-build applications
- Designers learn by wizarding

Disadvantages

- Simulations may misrepresent otherwise imperfect tech
- May simulate technologies that do not exist
- Wizards may need training and can be inconsistent
- Playing the wizard can be exhausting
- Some features (and limitations) can't be simulated
- May be inappropriate for certain venues (e.g., home)

Woz Testing in Studio Wed 2/15

Come with your group, prepared to setup and conduct a series of WOZ tests.

Members of your group will rotate out to serve as test participants for another group.

Rehearse ahead of time!

Start building your software! Next week you will show your “bare bones” functional prototype.

Controlled Experiments

What are the important concerns?

Controlled Experiments

Manipulate *independent variables (IV)*,
measure *dependent variables (DV)*.

Within or between-subjects design

Change IVs within or across subjects

Randomization, replication, blocking

Learning effects

Choice of measure and statistical tests

t-Test, ANOVA, Chi-squared χ^2 , Non-parametric

Subject population, number of participants

Experimental Desiderata

P-value: probability that results due to chance

Type I Error: accept spurious result

Bonferroni's principle: if you run enough
significance tests, you'll eventually get lucky

Type II Error: mistakenly reject result

Inappropriate measure or test?

Statistical vs. practical significance

N=1000, $p < 0.001$, avg dt = 0.012 sec.

Internal Validity

Internal validity: is a causal relation between two variables properly demonstrated?

Improper design or analysis: are there sufficient subjects? The right statistics?

Confounds: is there another factor at play?

Selection (bias): approp. subject population?

Experimenter bias: researcher actions

External Validity

External validity: do the results generalize to other situations of populations?

Subjects: are the subjects representative? do their aptitudes interact with the IVs?

Situation: time, location, lighting, duration

Ecological Validity

The degree to which the methods, materials and setting of the study approximate the real-life situation under investigation.

Flight simulator vs. flying a plane

Simulated community activity vs. open web





Evaluation Methods in HCI

Inspection (Walkthrough) Methods

Observation, User Studies

Experience Sampling

Interviews and Surveys

Usage Logging

Controlled Experimentation

Fieldwork, Ethnography

Mixed-Methods Approaches

**When should
you use which
method?**