

# Qualitative Evaluation

Why evaluation is crucial

Quickly debug prototypes by observing people use them

Methods reveal what a person is thinking about



# Qualitative Evaluation

## **Evaluating interfaces**

Lecture /slide deck produced by Saul Greenberg, University of Calgary, Canada



# Overview

---

Why evaluation is crucial

Quickly debug prototypes by observing people use them

Methods reveal what a person is thinking about

# Why bother?

---

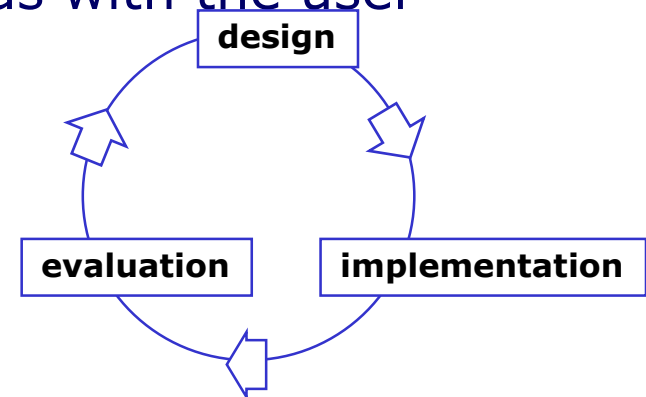
Tied to the usability engineering lifecycle

## Pre-design

- investing in new expensive system requires proof of viability

## Initial design stages

- develop and evaluate initial design ideas with the user



# Why bother?

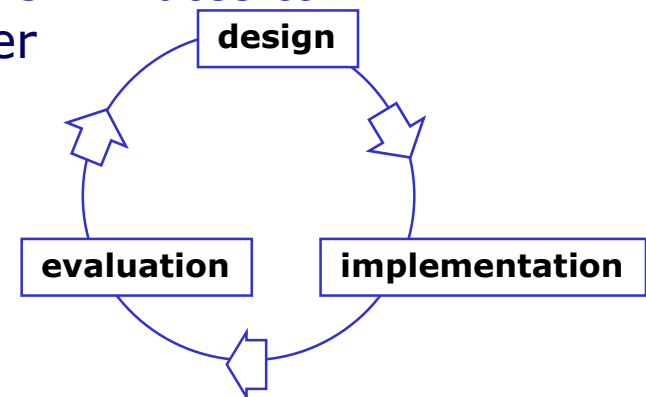
---

## Iterative design

- does system behavior match the user's task requirements?
- are there specific problems with the design?
- what solutions work?

## Acceptance testing

- verify that system meets expected user performance criteria
  - 80% of 1st time customers will take 1-3 minutes to withdraw \$50 from the automatic teller



# Naturalistic approach

---

## Observation occurs in realistic setting

- real life
- Provides useful, realistic data
- More likely to generalize (though do not assume it will)

## Problems

- hard to arrange and do
- time consuming



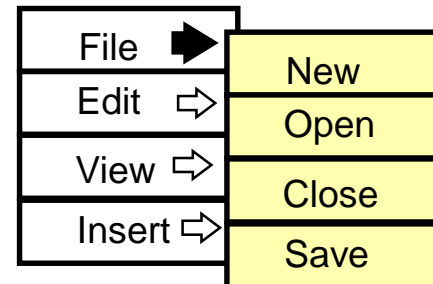
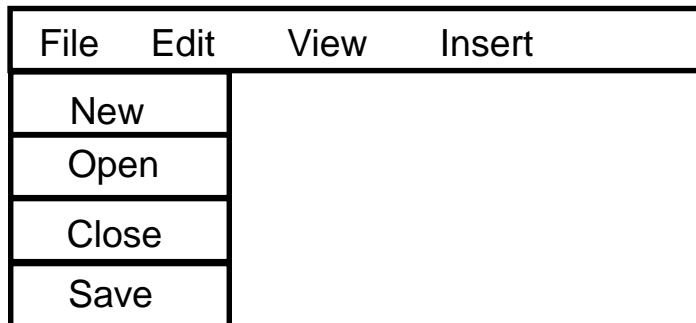
# Experimental approach

---

Experimenter controls all environmental factors

- study relations by manipulating *independent* variables
- observe effect on one or more *dependent* variables
- nothing else changes

*There is no difference in user performance (**time** and **error rate**) when selecting an item from a **pull down** or a **pull right** menu of 4 items*



# Validity

---

## *External validity*

- confidence that results applies to real situations
- usually good in natural settings

## *Internal validity*

- confidence in our explanation of experimental results
- usually good in experimental settings

## Trade-off: Natural vs Experimental

- precision and direct control over experimental design  
*versus*
- desire for maximum generalizability in real life situations



# **Usability engineering approach**

---

Observe people using systems in simulated settings

- people brought in to artificial setting that simulates aspects of real world setting
- people given specific tasks to do
- observations / measures made as people do their tasks
- look for problem areas / successes
- good for uncovering 'big effects'



# Usability engineering approach

---

Is the test result relevant to the usability of real products in real use outside of lab?

## Problems

- non-typical users tested
- non-typical tasks
- different physical environment
- different social context
  - motivation towards experimenter vs motivation towards boss

## Partial Solution

- use real users
- task-centered system design tasks
- environment similar to real situation



# Usability engineering approach

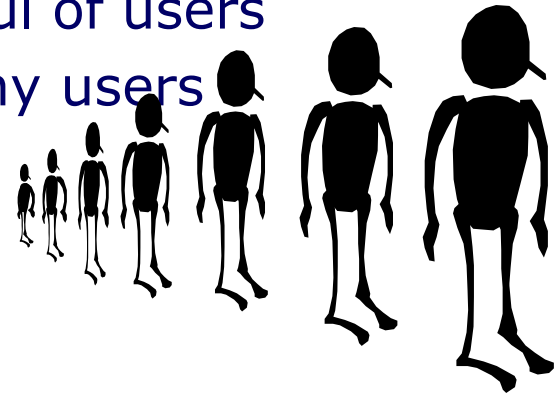
---

How many users should you observe?

- observing many users is expensive
- *but* individual differences matter
  - best user 10x faster than slowest
  - best 25% of users  $\sim$ 2x faster than slowest 25%

partial solution

- reasonable number of users tested
- reasonable range of users
- big problems usually detected with handful of users
- small problems / fine measures need many users



# Discount usability evaluation

---

## Low cost methods to gather usability problems

- approximate: capture most large and many minor problems

## How?

- qualitative:
  - observe user interactions
  - gather user explanations and opinions
  - produces a description, usually in non-numeric terms
  - anecdotes, transcripts, problem areas, critical incidents...
- quantitative
  - count, log, measure something of interest in user actions
  - speed, error rate, counts of activities,

# Discount usability evaluation

---

## Methods

- inspection
- extracting the conceptual model
- direct observation
  - think-aloud
  - constructive interaction
- query techniques (interviews and questionnaires)
- continuous evaluation (user feedback and field studies)

# Inspection

---

Designer tries the system (or prototype)

- does the system “feel right”?
- benefits
  - can catch some major problems in early versions
- problems
  - not reliable as completely subjective
  - not valid as introspector is a non-typical user
  - intuitions and introspection are often wrong

Inspection methods help

- task centered walkthroughs
- heuristic evaluation



# Conceptual model extraction

---

How?

- show the user static images of
  - the prototype or screens during use
- ask the user explain
  - the function of each screen element
  - how they would perform a particular task and why they think that

What?

- **Initial conceptual model**
  - how person perceives a screen the very first time it is viewed
- **Formative conceptual model**
  - How person perceives a screen after its been used for a while

Value?

- good for eliciting people's understanding before & after use
- poor for examining system exploration and learning

# Direct observations

---

Evaluator observes users interacting with system

- in lab:
  - user asked to complete a set of pre-determined tasks
- in field:
  - user goes through normal duties

## Value

- excellent at identifying gross design/interface problems
- validity depends on how controlled/contrived the situation is



# **Simple observation method**

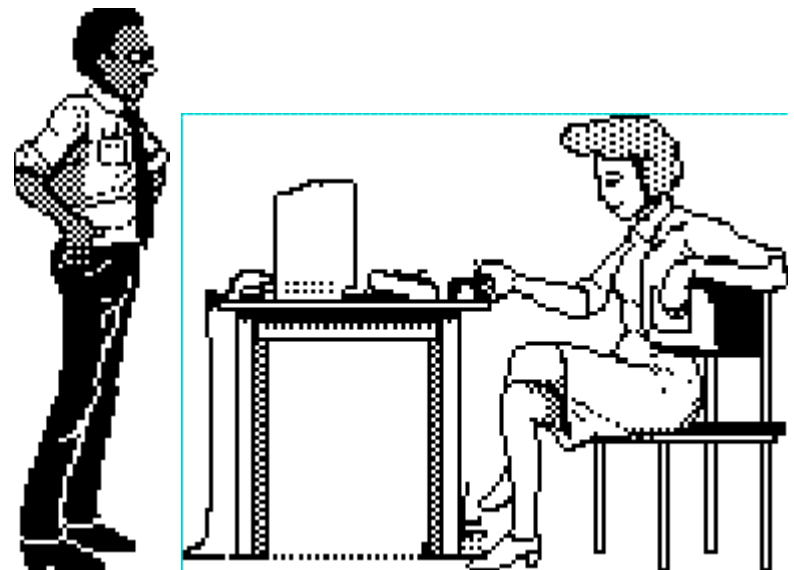
---

User is given the task

Evaluator just watches the user

## **Problem**

- does not give insight into the user's decision process or attitude

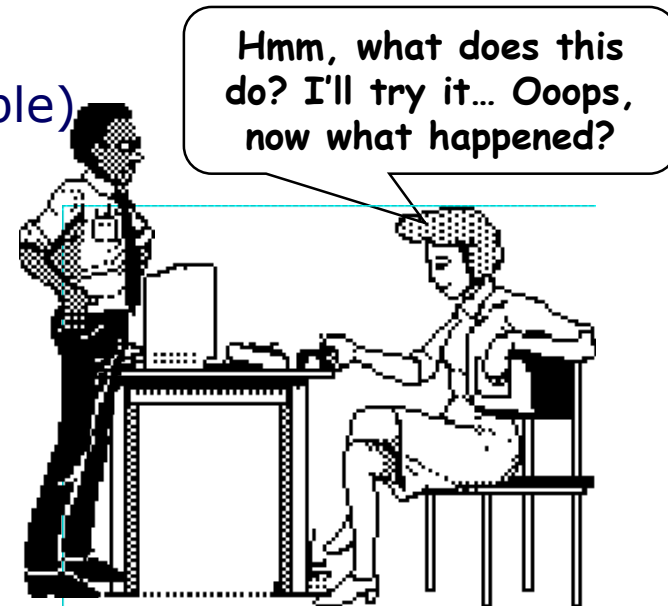


# Think aloud method

---

Users speak their thoughts while doing the task

- what they are trying to do
- why they took an action
- how they interpret what the system did
- gives insight into what the user is thinking
- most widely used evaluation method in industry
  - may alter the way users do the task
  - unnatural (awkward and uncomfortable)
  - hard to talk if they are concentrating



# Example Usability Test

---

Drawing some cats!

<https://www.youtube.com/watch?v=9wQkLthhHKA>



# Constructive interaction method

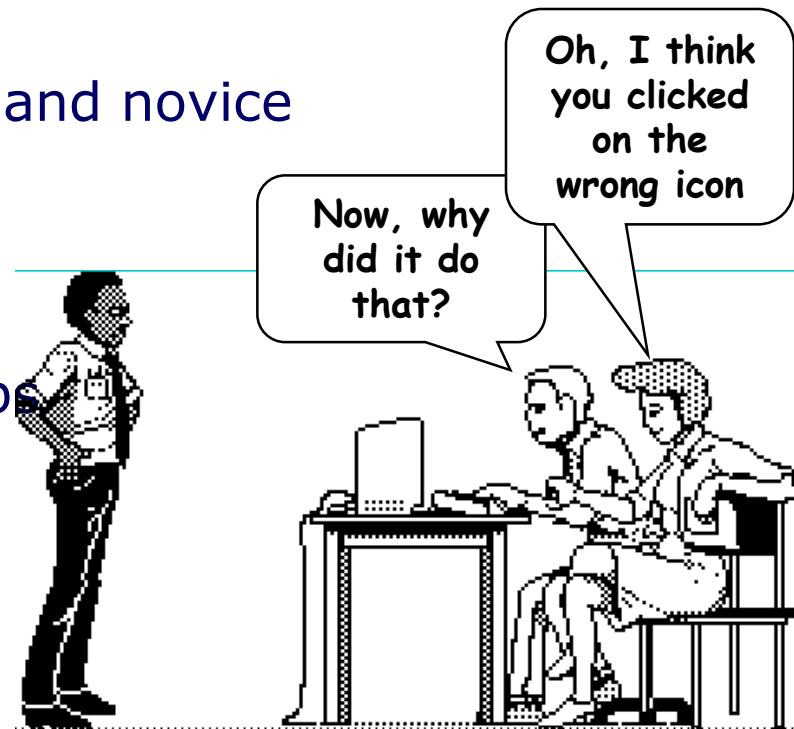
---

Two people work together on a task

- monitor their normal conversations
- removes awkwardness of think-aloud

Co-discovery learning

- use semi-knowledgeable “coach” and novice
- only novice uses the interface
  - novice ask questions
  - coach responds
- gives insights into two user groups

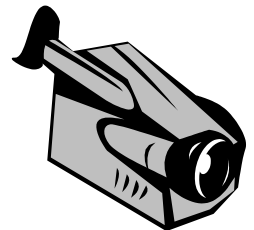


# Recording observations

---

How do we record user actions for later analysis?

- otherwise risk forgetting, missing, or misinterpreting events
- paper and pencil
  - primitive but cheap
  - observer records events, comments, and interpretation
  - hard to get detail (writing is slow)
  - 2<sup>nd</sup> observer helps...
- audio recording
  - good for recording think aloud talk
  - hard to tie into on-screen user actions
- video recording
  - can see and hear what a user is doing
  - one camera for screen, rear view mirror useful...
  - initially intrusive

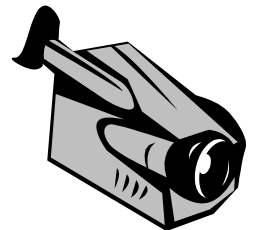


# Recording observations

---

There is software for this:

- Silverback (<http://silverbackapp.com/>)
- Morae (<http://www.techsmith.com/morae.html>)
  - This one is EXPENSIVE!
- Camtasia
- More??



# Interviews

---

## Good for pursuing specific issues

- vary questions to suit the context
- probe more deeply on interesting issues as they arise
- good for exploratory studies via open-ended questioning
- often leads to specific constructive suggestions

## Problems:

- accounts are subjective
- time consuming
- evaluator can easily bias the interview
- prone to rationalization of events/thoughts by
  - user's reconstruction may be wrong



# How to Interview

---

## Plan a set of central questions

- a few good questions gets things started
  - avoid leading questions
- focuses the interview
- could be based on results of user observations

## Let user responses lead follow-up questions

- follow interesting leads vs bulldozing through question list



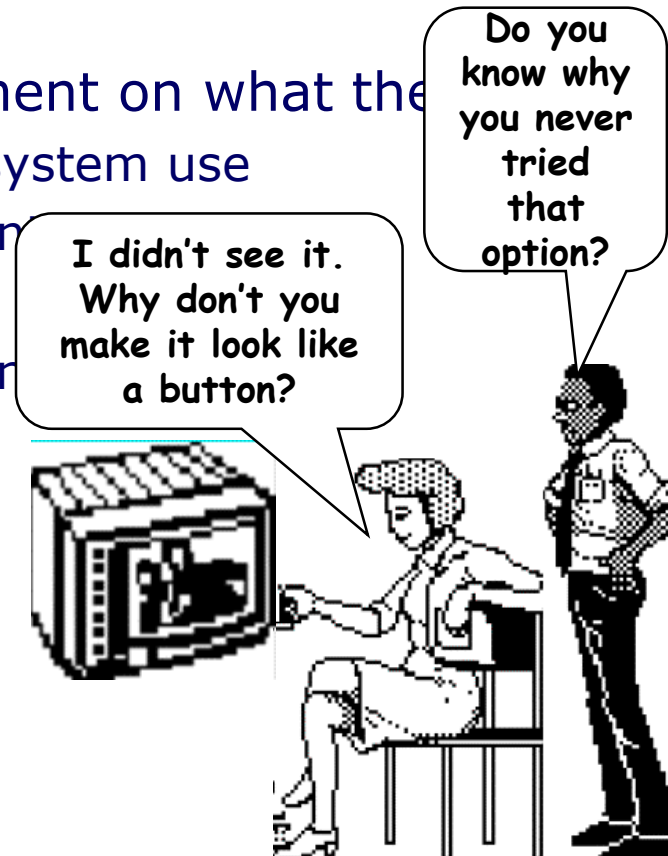


# Retrospective testing interviews

---

## Post-observation interview to

- perform an observational test
- create a video record of it
- have users view the video and comment on what they saw
  - clarify events that occurred during system use
  - excellent for grounding a post-test in actual events
  - avoids erroneous reconstruction
  - users often offer concrete suggestions

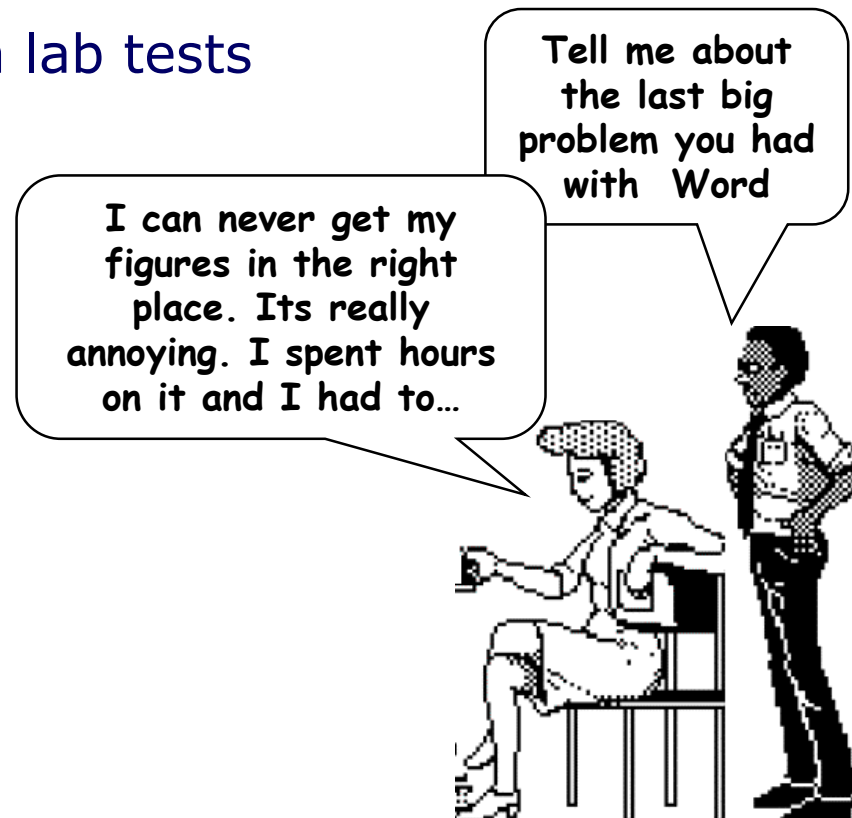


# Critical incidence interviews

---

People talk about incidents that stood out

- usually discuss extremely annoying problems with fervor
- not representative, but important to them
- often raises issues not seen in lab tests



# Questionnaires and Surveys

---

## Questionnaires / Surveys

- preparation “expensive,” but administration cheap
  - can reach a wide subject group (e.g. mail)
- does not require presence of evaluator
- results can be quantified

But

- only as good as the questions asked



# Questionnaires and Surveys

---

## How

- establish the purpose of the questionnaire
  - what information is sought?
  - how would you analyze the results?
  - what would you do with your analysis?
- do not ask questions whose answers you will not use!
- determine the audience you want to reach
- determine how would you will deliver / collect the questionnaire
  - on-line for computer users
  - web site with forms
  - surface mail
    - pre-addressed reply envelope gives far better response

# Continuous Evaluation

---

## Monitor systems in actual use

- usually late stages of development
  - ie beta releases, delivered system
- fix problems in next release

## User feedback via gripe lines

- users can provide feedback to designers while using the system
  - help desks
  - bulletin boards
  - email
  - built-in gripe facility
- best combined with trouble-shooting facility
  - users always get a response (solution?) to their gripes



# What you now know

---

Debug designs by observing how people use them

- quickly exposes successes and problems
- specific methods reveal what a person is thinking
- but naturalistic vs laboratory evaluations is a tradeoff

Methods include

- conceptual model extraction
- direct observation
  - think-aloud
  - constructive interaction
- query via interviews, retrospective testing and questionnaires
- continuous evaluation via user feedback and field studies

# **You know now**

---

Why evaluation is crucial

Quickly debug prototypes by observing people use them

Methods reveal what a person is thinking about

# Primary Sources

---

This slide deck is partly based on concepts as taught by:

- Nielsen, J. (1993) Usability Engineering, Chapter 6: Usability testing
- Gomoll, Kathleen & Nicol, Anne (1990) User Observation: Guidelines for Apple Developers, Apple Inc., January
- Dumas, J.S. and Redish, J.C. A Practical Guide to Usability Testing. Revised Edition. (1999)
- Gould, J. (1988) How to design usable systems. In Readings in Human Computer Interaction: Towards the Year 2000 (2nd Edition). Baecker, R., Grudin, J., Buxton, W., and Greenberg, S. (1995). Morgan-Kaufmann.