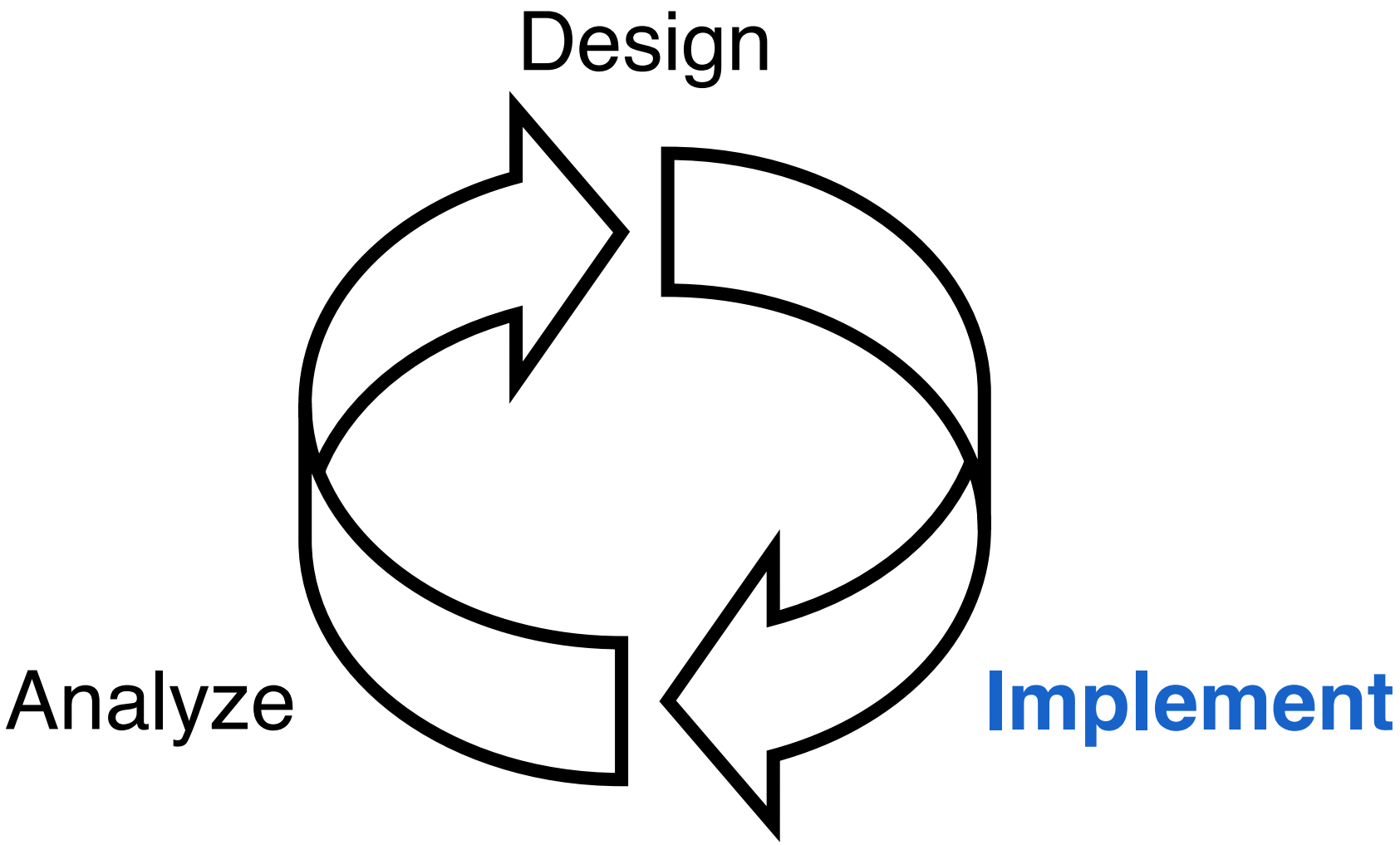
Prototyping and Evaluation of Mobile Systems

Mensch-Maschine-Interaktion 2, WS 2010/2011

Michael Rohs
michael.rohs@ifi.lmu.de
MHCI Lab, LMU München

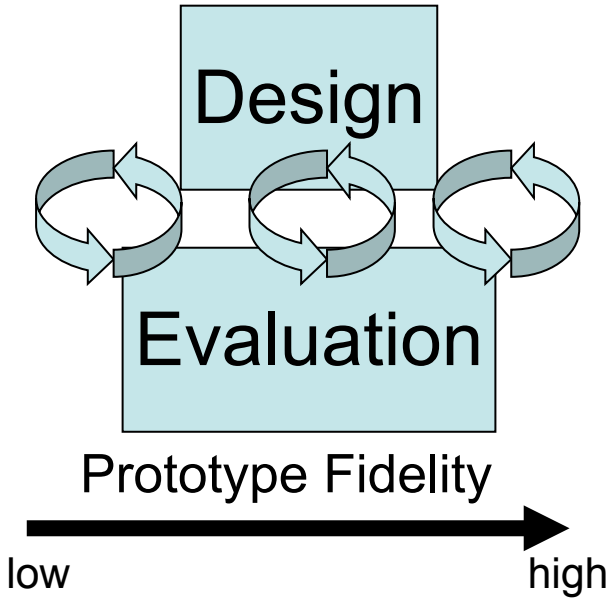
DIA Cycle: How to realize design ideas?



Prototyping

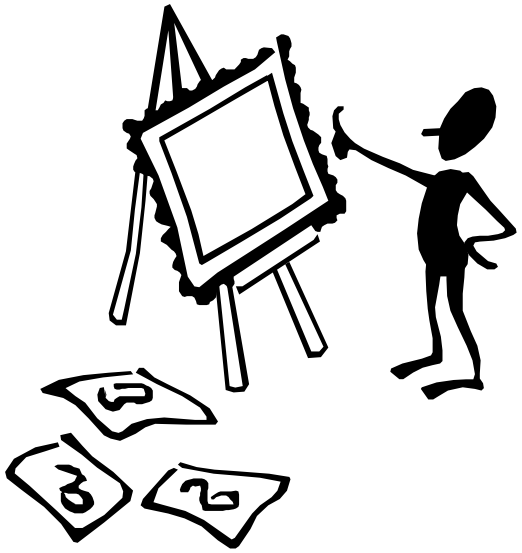
From Ideas to Implementation: Prototyping

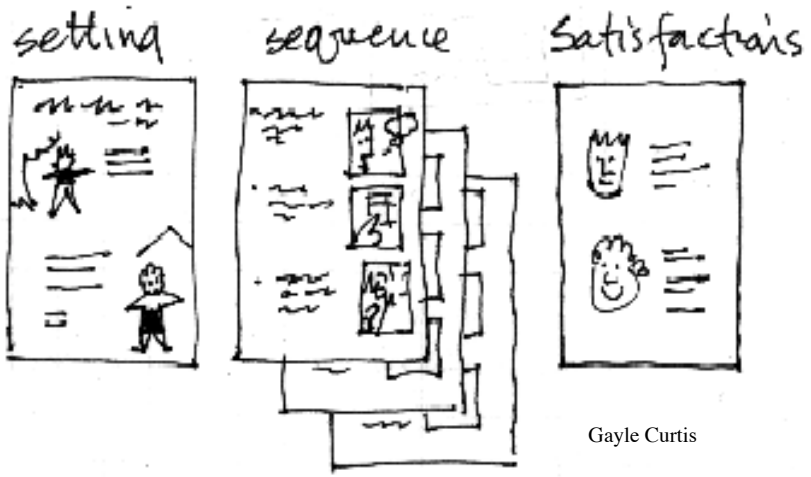
- Building a scaled-down version of an interactive system to collect information to guide its further design
  - Invaluable for iterative design
- Get early feedback on emerging designs
  - After initial requirements analysis, scenarios
- Continuous input for design decisions
  - During all design phases
- Prototype appropriate for
  - Audience
  - Design phase
  - Design question



Low-Fidelity Paper Prototypes

- First prototype, quick and cheap
- Paper and pencil mockup of user interface
  - Rough sketches of the main screens and dialogs
  - Textual description of interface functions and relationships between screens
- Goals
  - Brainstorming
  - Expert review of interaction flow
  - First user feedback
  - User tests





Gayle Curtis

Paper / Post-it Prototype Process

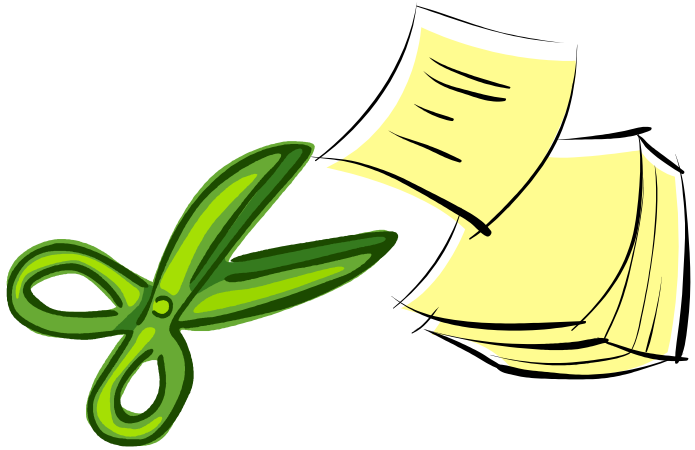
Collaboratively creating the prototypes

Reviewing the prototypes

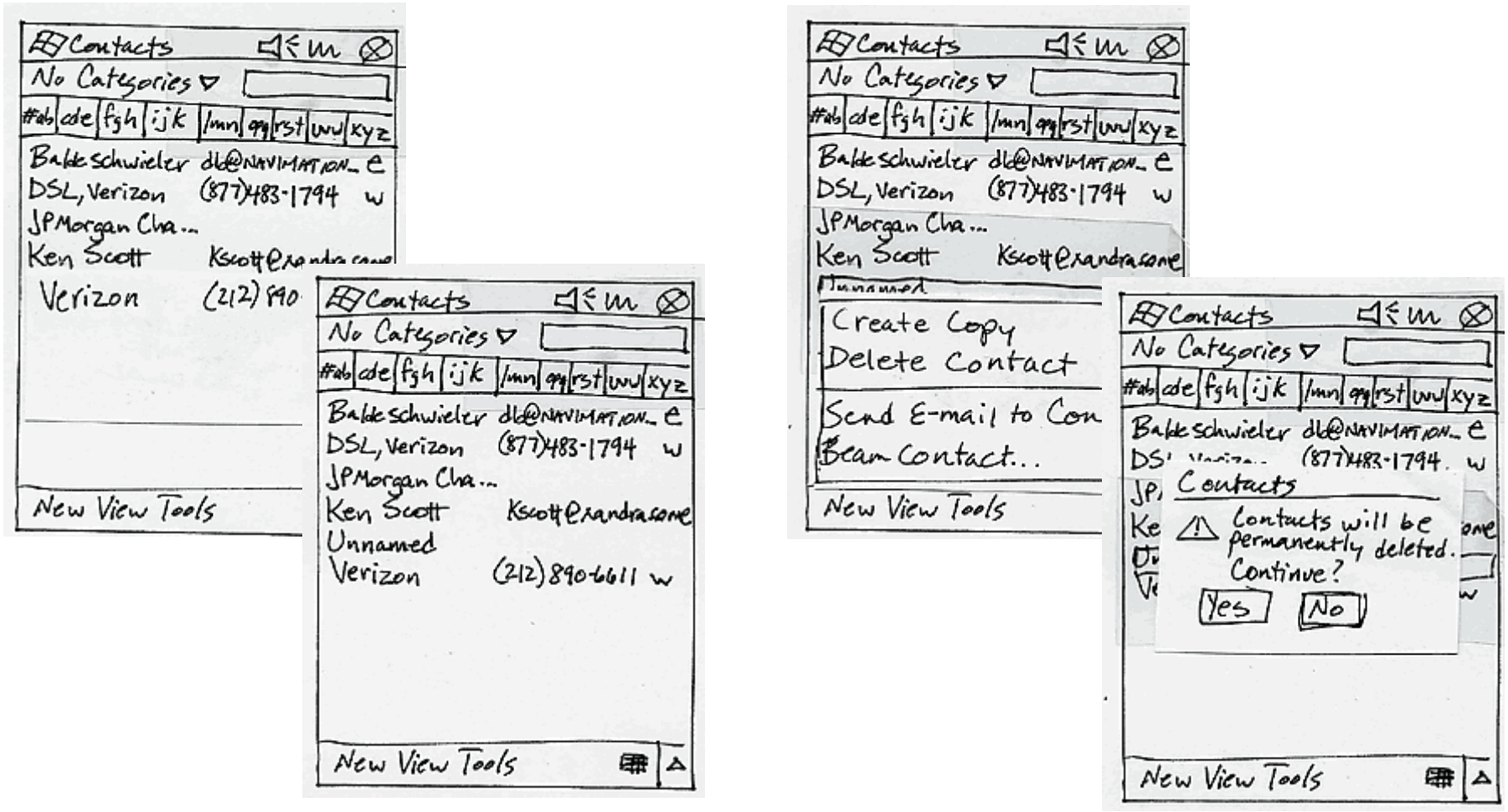
Source: http://www.pocketpcmag.com/_archives/may03/e_prototyping.asp

Building a Low-Fidelity Prototype

- Assemble material
  - Paper: large heavy paper for designs, cards for note taking
  - Adhesives: tape, glue sticks, correction tape
  - Markers: colored pens and pencils, highlighters, fine liners
  - Scissors
  - Post-it notes
  - Transparent sheets for user input

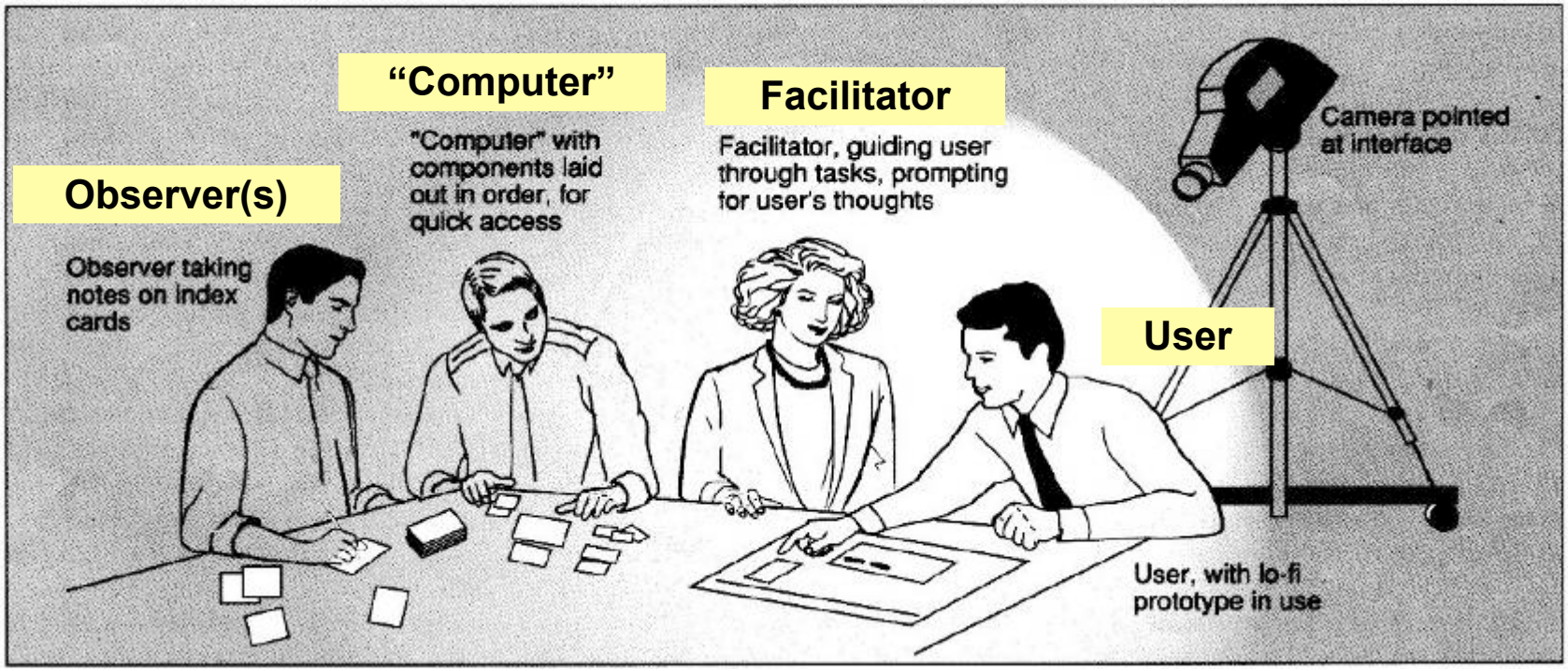


Paper Prototype Examples



Source: http://www.pocketpcmag.com/_archives/may03/e_prototyping.asp

Low-Fidelity User Testing

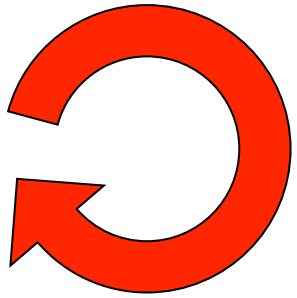


Marc Rettig: Prototyping for Tiny Fingers

- Select users
- Prepare test scenarios, drawn from task analysis
  - familiar data, realistic tasks
- Practice
  - team members know their roles, no “computer” delays

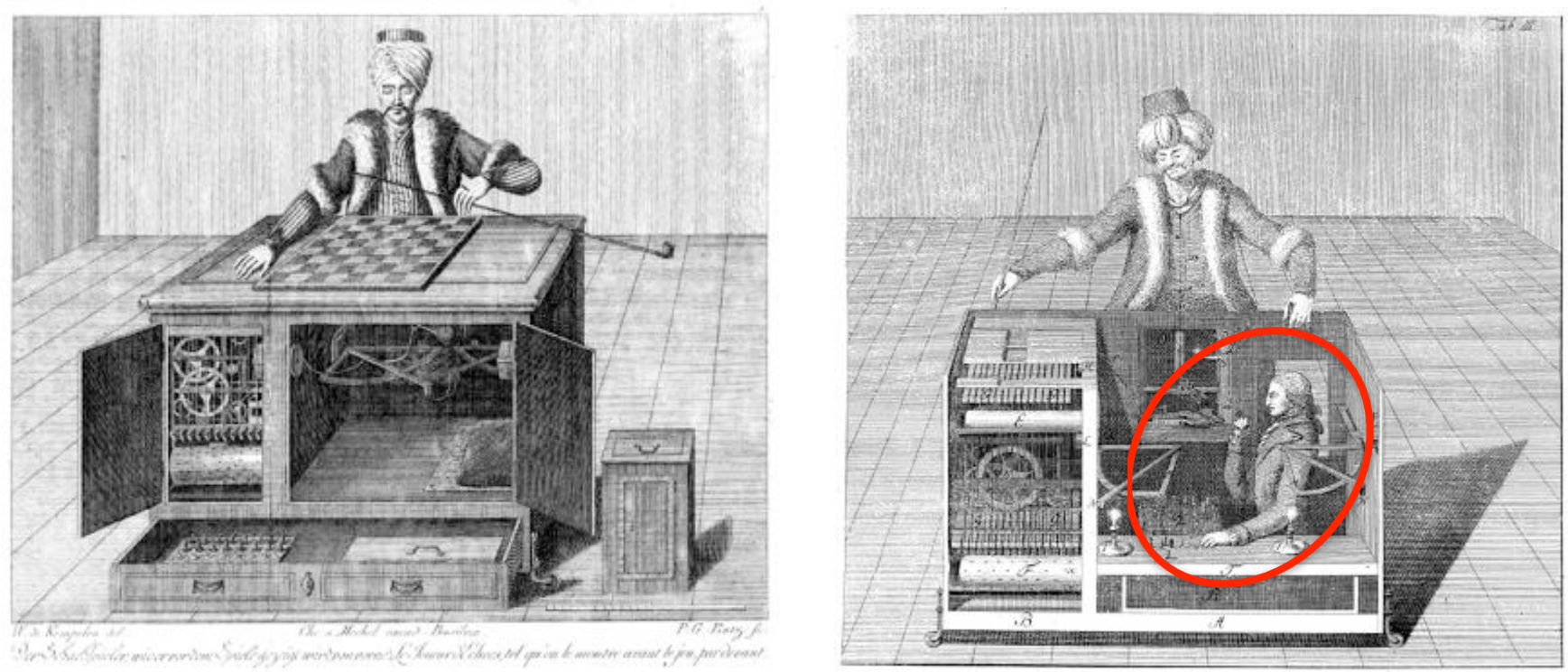
Low-Fidelity Prototype Revision

- Evaluation of test results
  - Arrange paper prototype on table
  - Pile note cards next to component
- Summarize and prioritize problems
  - Written report on findings
- Prototype refinement
  - Agenda for meeting to discuss design changes
  - Attach post-it notes with changes to each component



Wizard of Oz …

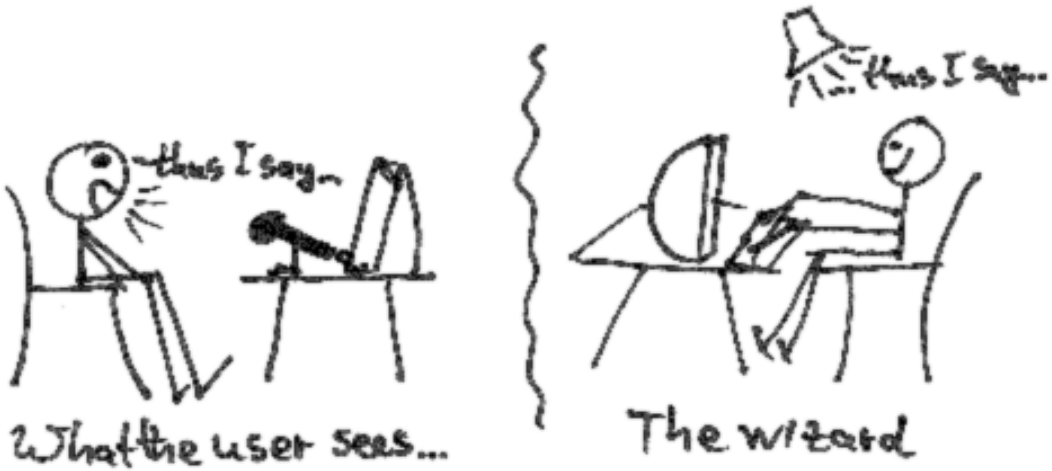
- The first “Chess Computer”
  - “In 1769, Hungarian nobleman Wolfgang von Kempelen astonished Europe by building a mechanical chess-playing automaton that defeated nearly every opponent it faced.”



Source: http://collabor.f4.fhtw-berlin.de:8888/mmgestalt07s/topics/Beispiel/

… Wizard of Oz

- Method for testing a non-existing system
- Human “wizard” simulates system responses
  - Interacts with user via a simulated software user interface
- Useful for adding complex vertical functionality
  - Speech and gesture recognition, language translation

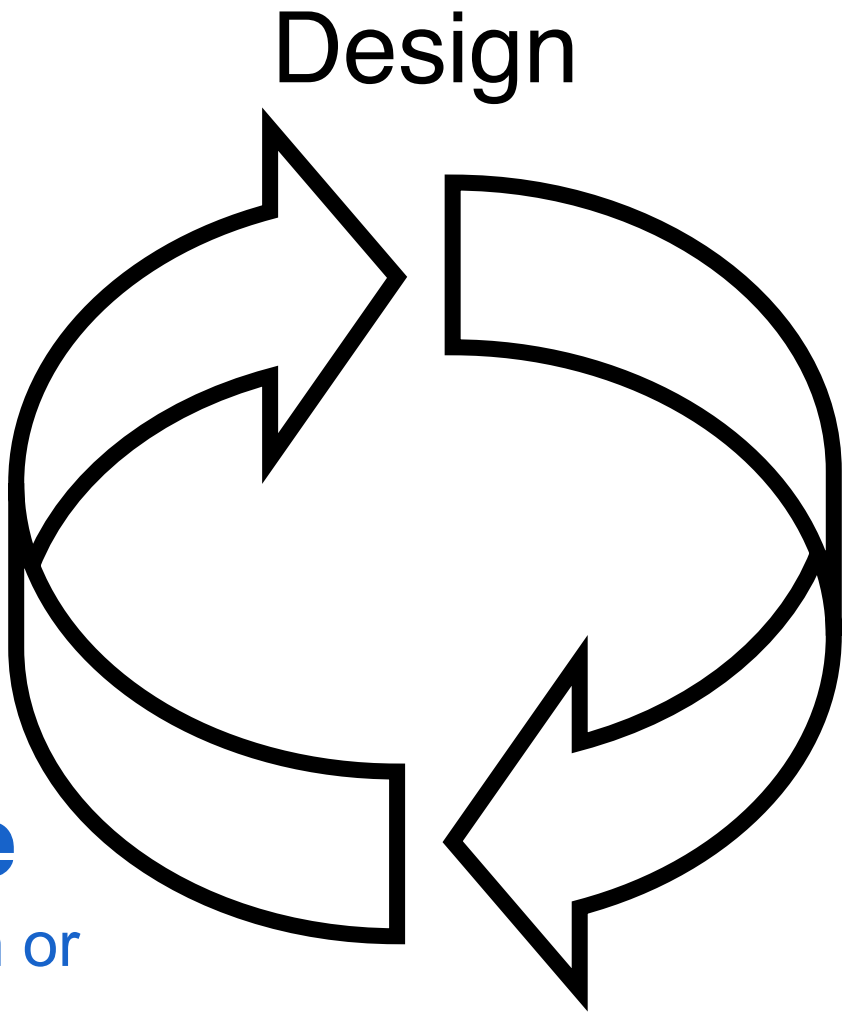


Summary

- Prototyping enables rapid feedback on design ideas
- Different kinds for different purposes and design stages
- Choose appropriate prototype for question to answer
- Low-fidelity vs. medium-fidelity vs. high-fidelity
- Many approaches, methods, and tools

Evaluation

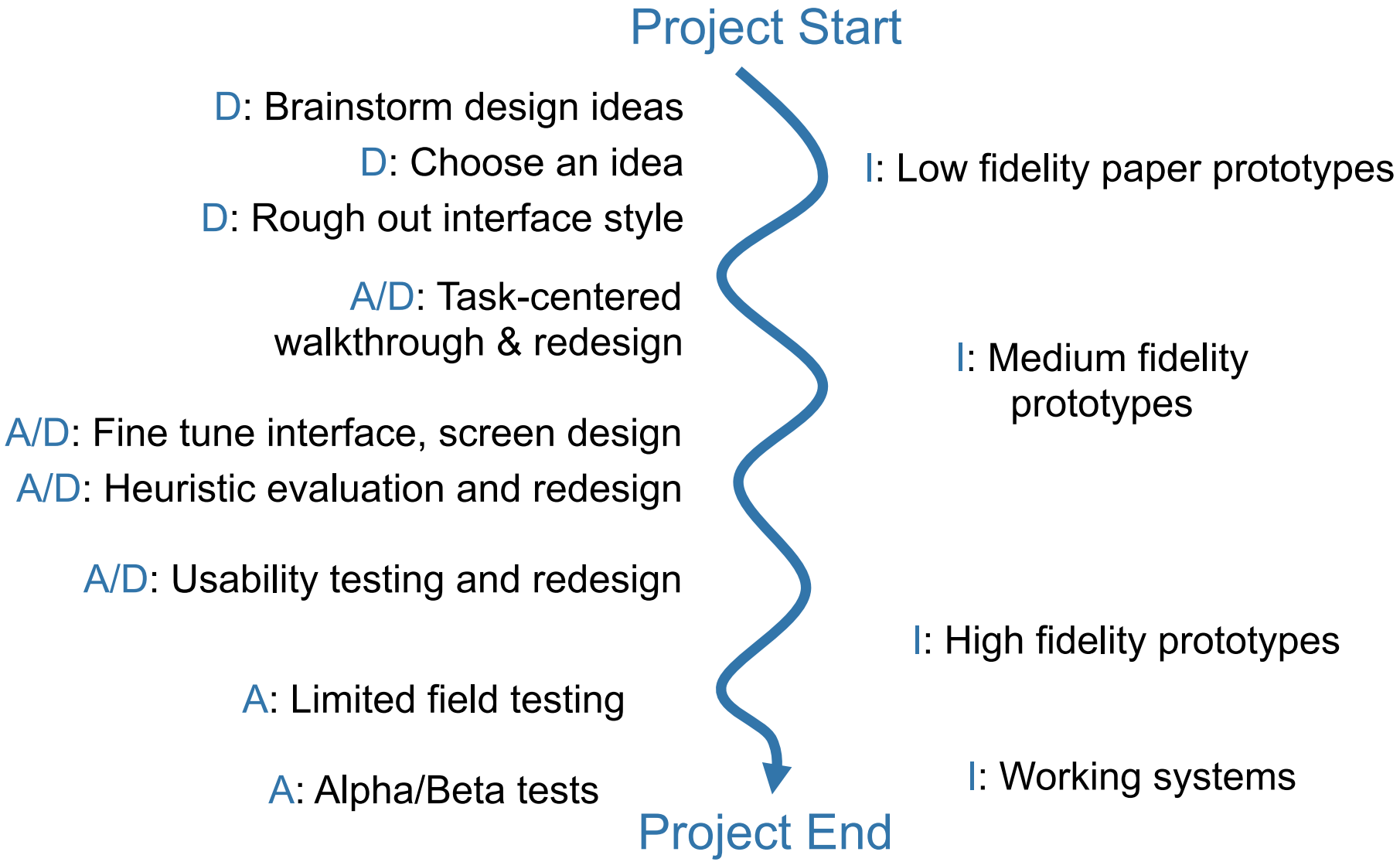
DIA Cycle: When to evaluate?



Analyze

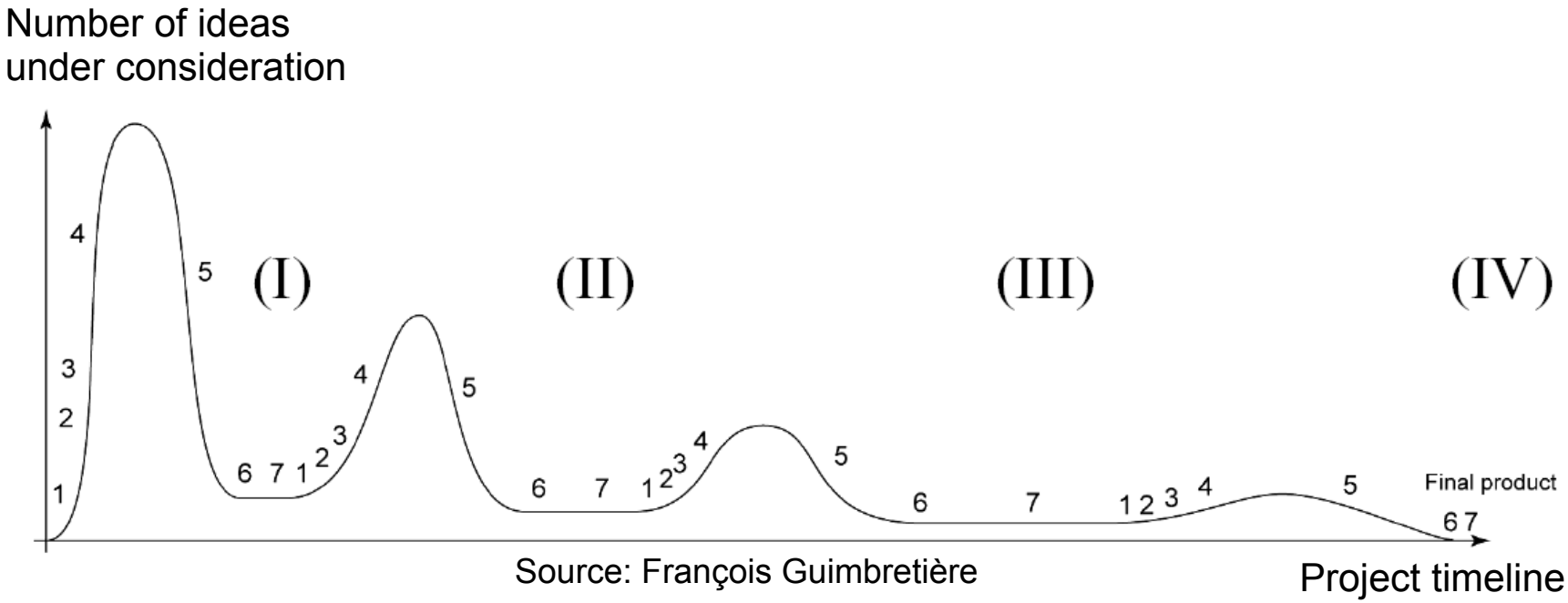
Evaluate with or without users

Design – Implement – Analyze



Evaluation Depending on Project Phase

- Walk-throughs and paper based interface (I)
- Simulation of the interface and Wizard of Oz (II)
- Larger and larger group of users using the real interface (III)
- Product is shipping (IV)



Source: François Guimbretière

Why evaluate?

- To ensure that system matches user needs or predefined usability goals (usability engineering)

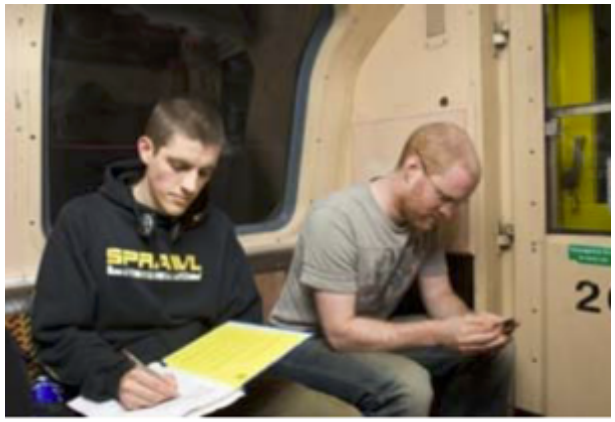
- Judge system functionality
  - Does it facilitate users' tasks?
  - Does it offer the right features, easy to reach?
- Judge effects on users
  - How easy is the system to learn and use?
  - How do users feel about the system? → "Joy of use"?
  - Are there areas that overload users?
- Discover specific problems
  - Do unexpected / confusing situations come up?

Where to evaluate: Laboratory

Source: http://wwwswt.informatik.uni-rostock.de/deutsch/Mitarbeiter/michael/lehre/Usab_WS2002/Jan/vortrag_html.htm

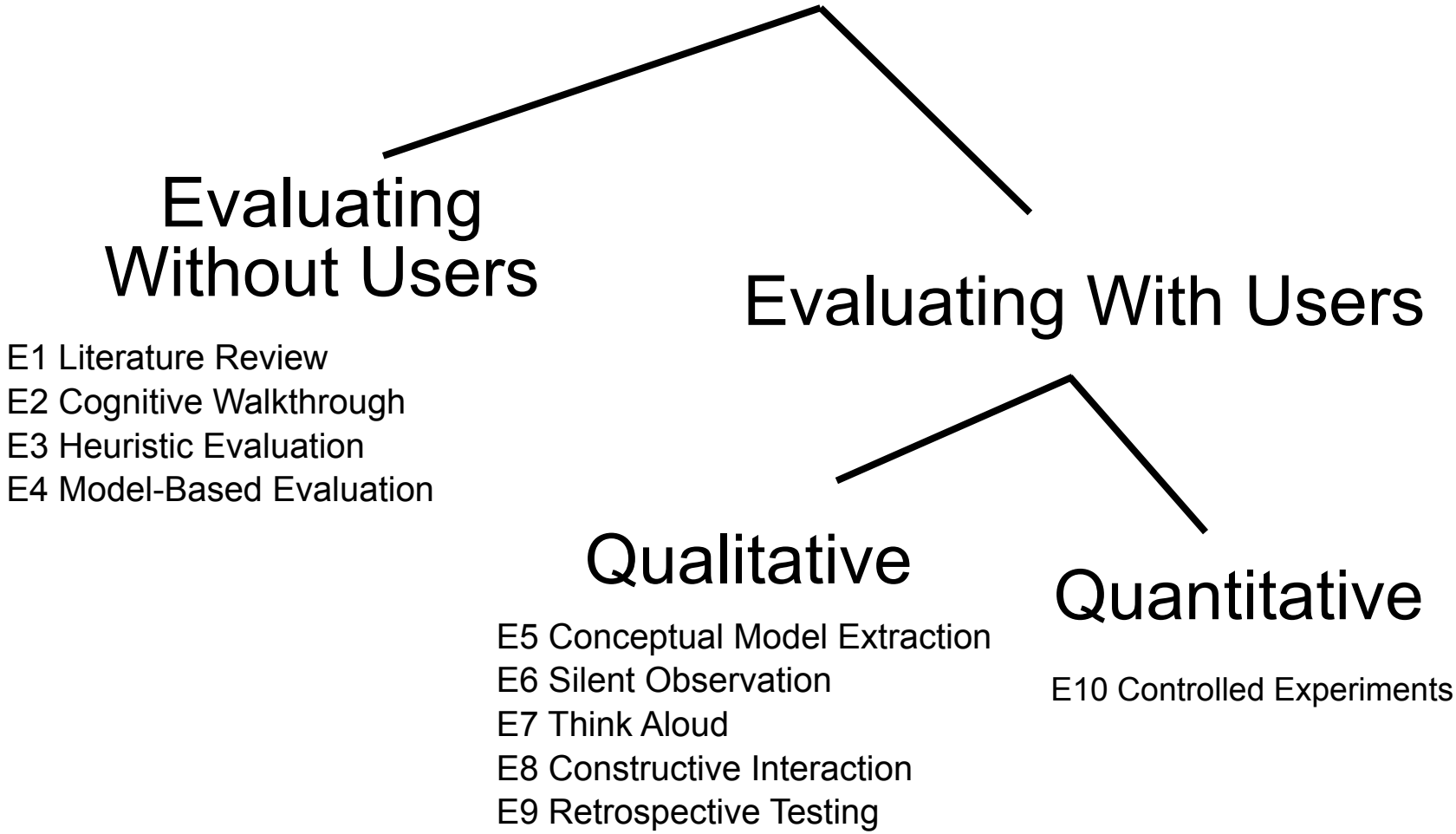
- With or without users
+ Equipment (audio / video, see-through mirrors, special computers), no disruptions, quiet
– Natural environment missing (shelves, wall calendar, streets, people…); unnatural situation (relevance?)
- Only place possible if real use dangerous, remote (ISS…), or controlled situation needed

Where to evaluate: In the field


- Studies in the users' natural environment
- Advantages
  - + Situations (location and context!) and behavior more natural
  - + More realistic (also because of disruptions)
  - + Better suited to long-term studies
- Disadvantages
  - – Noise, task interruptions
  - – Will still feel like a test situation

Evaluation Techniques



+ Interviews, questionnaires,...

E1: Literature Review

- Many research results about user interface design have been published
- Idea: Search literature for evidence for (or against) aspects of your design

\+ Saves own experiments

– Results only carry over reliably if context (users, assumptions) is very similar

E2: Cognitive Walkthrough

- Analytical method for early design or existing systems
  - Without users
- Expert evaluator = designer or cognitive psychologist
- Goal: Judge learnability and ease of use
  - Does system help user to get from goals to intentions and actions?
- Step through each action and ask
  - Is the effect of the action the same as the user's goal at that point?
  - Will users see that the action is available?
  - Once users find the action, will they know it is the right one?
  - After the action is taken, will users understand the feedback?

E2: Cognitive Walkthrough

- What you need
  - Interface description (prototype of the system)
  - Task description
    - Example: Program the video to time-record a program starting at 18:00 and finishing at 19:15 on channel 4 on January 26, 2011
  - List of interface actions to complete the task
  - User profile
- Doing the actual walkthrough
  - Analyze process of performing the actions using above questions
- Written questions capture psychological knowledge and guide the tester

E3: Heuristic Evaluation

- Variant of Cognitive Walkthrough
- Choose usability heuristics
  - (general usability principles, e.g., Nielsen's 10 Usability Principles)
- Step through tasks and check whether guidelines are followed
- Severity rating for each problem (Nielsen)
  - 0 = I don't agree this is a problem at all
  - 1 = cosmetic problem
  - 2 = minor usability problem, low priority to fix
  - 3 = major usability problem, high priority to fix
  - 4 = usability catastrophe, imperative to fix before release

\+ Quick and cheap

– Subjective (better done by several independent evaluators)

- See also: www.useit.com/papers/heuristic

10 Usability Principles (Jakob Nielsen)

1. Keep the interface simple!
2. Speak the user's language!
3. Minimize the user's memory load!
4. Be consistent and predictable!
5. Provide feedback!
6. Design clear exits and closed dialogs!
7. Offer shortcuts for experts!
8. Help to recover from errors, offer Undo!
9. Prevent errors!
10. Include help and documentation!

8 Golden Rules of Interface Design (Ben Shneiderman)

1.	Strive for consistency	Sequences, terminology, layout
2.	Cater to universal usability	Diverse users, experience, needs
3.	Offer informative feedback	Direct manipulation, subtle feedback
4.	Design dialogs to yield closure	Grouping of related interactions
5.	Prevent errors	Gray out items, numeric input fields
6.	Permit easy reversal of action	Allow undo, encourage exploration
7.	Support internal locus of control	Minimize surprise, users as initiators rather then responders of actions
8.	Reduce short-term memory load	7 ±2, reduce abbreviation

User Interface Guidelines

- Concrete guidelines for look-and-feel and behavior
  - Visual appearance, e.g., icon design
  - Purpose of user interface elements
  - Layout of user interface elements
  - Behavior, conventions of system features
- Android User Interface Guidelines
  - http://developer.android.com/guide/practices/ui_guidelines/index.html
- iOS Human Interface Guidelines
  - http://developer.apple.com/library/ios/documentation/userexperience/conceptual/mobilehig/MobileHIG.pdf
  - Aesthetic integrity, consistency, direct manipulation, feedback, metaphors, user control, …

E4: Model-Based Evaluation

- Several theoretical models exist that offer a framework for design and evaluation
- Examples
  - GOMS (= goals, operators, methods, selection rules)
  - KLM (= keystroke-level model)
  - Design Rationale (history of design decisions with reasons and alternatives)
  - Design Patterns

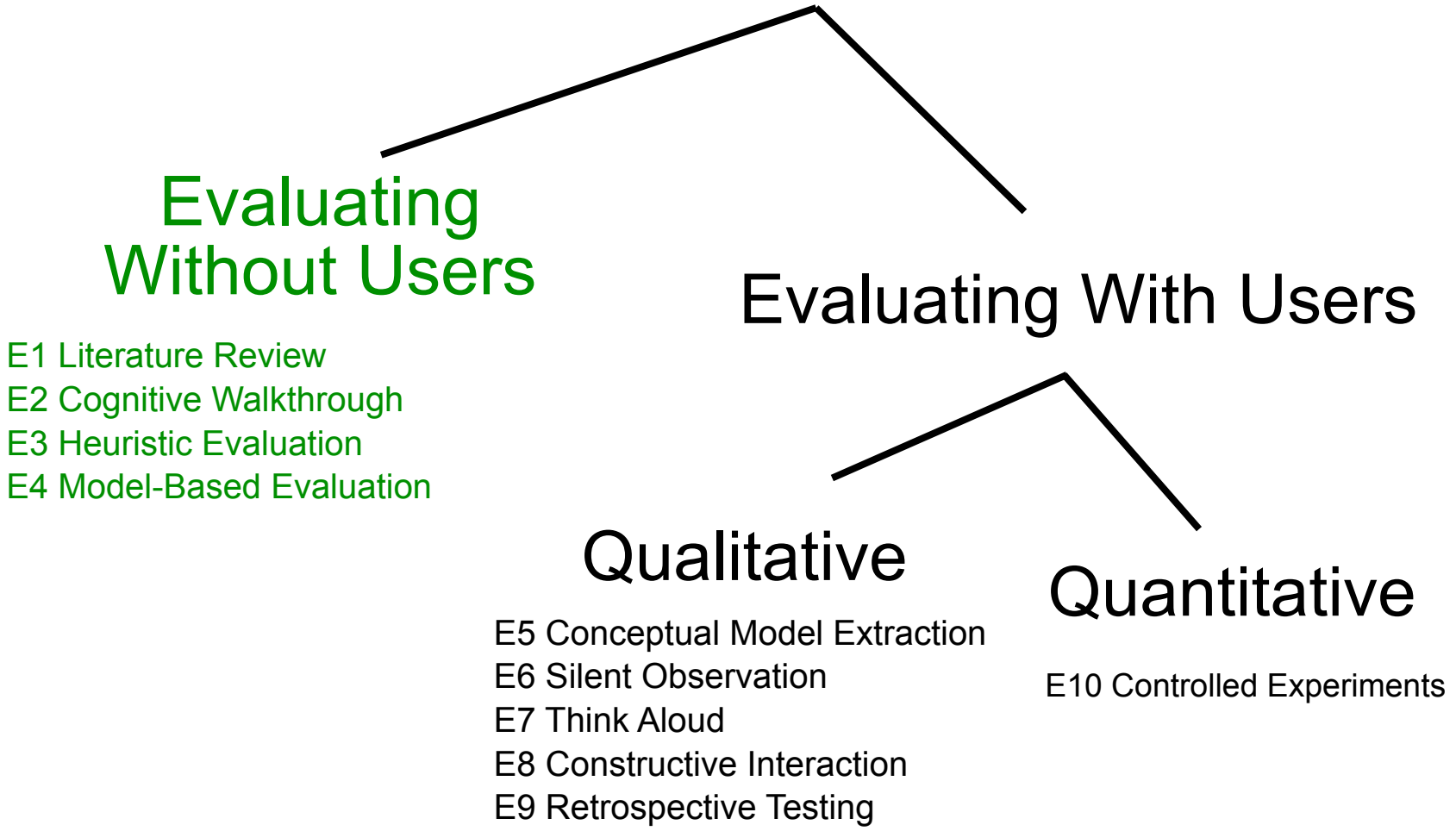
GOMS Analysis

- **G**oals, **O**perators, **M**ethods, **S**election rules
  - Card, Moran, Newell: The Psychology of HCI, 1983
- Model of task execution with a given system
  - To estimate execution times, mental effort, and learnability before a system is built
- Model of user knowledge necessary to do task execution
  - Procedural knowledge (skills) about executing operators
  - Declarative knowledge about goal structures
- Expert users performing routine tasks
  - Not creative tasks or problem-solving

GOMS: Components

- Goals describe user's end goals
  - E.g., "copyedit manuscript"
  - Leads to hierarchy of subgoals
- Operators are elementary user actions (mental or external)
  - Key presses, menu selection, drag & drop, speech commands,…
  - Assign context-independent duration (in ms)
- Methods are procedures to reach a goal
  - Consist of subgoals and/or operators
  - E.g., delete some text
- Selection rules
  - Which method to use for a (sub)goal
  - Selection depends on contents of STM state

Evaluation Techniques



+ Interviews, questionnaires,...

Evaluating With Users

- E1–E4 evaluate designs without the user
- As soon as implementations (prototypes) exist they should also be tested with users, using the following methods

Dealing with Test Users

- Tests are uncomfortable for the tester
  - Pressure to perform, mistakes, competitive thinking
- So treat testers with respect at all times!
  - Before, during, and after the test

Participatory Design

- Involve user as part of design team throughout entire software process
- Originated in Scandinavia where it is the law
- Techniques for team communication
  - Brainstorming, storyboarding, workshops, interviews, role plays, paper prototypes

E5: Conceptual Model Extraction

- Designer shows user prototype or screen shots
- User tries to explain elements and their function

\+ Good to understand naïve user's conceptual model of the system

– Bad to understand how the system is learned over time

E5: Conceptual Model Extraction Example

What do these icons mean (in a digital camera)?

Orderly stack

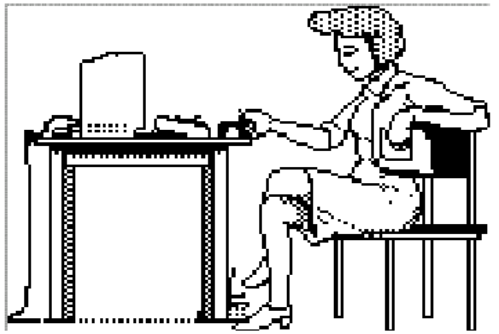
“Messy” stack

Taking pictures of skycrapers?

Viewing taken pictures!

Source: Jones and Marseden: Mobile Interaction Design

E6: Silent Observation



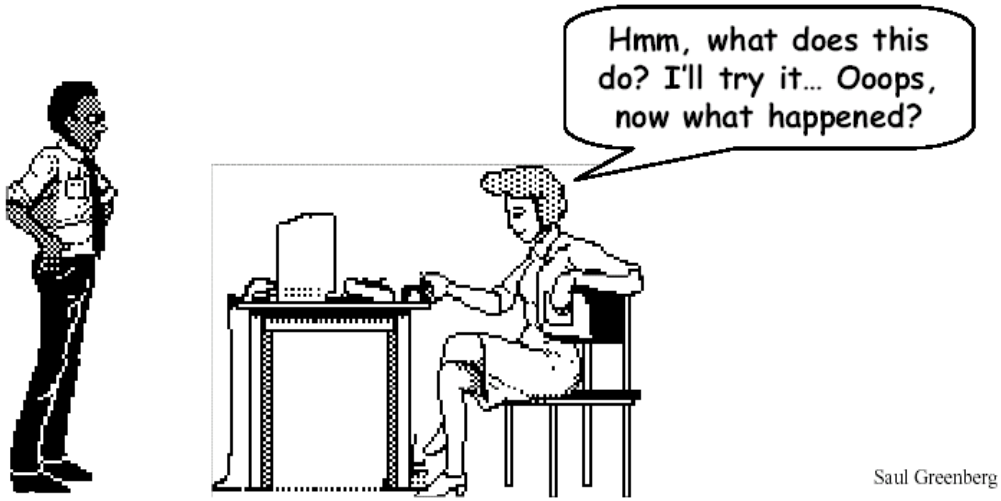
Saul Greenberg

- Designer watches user in lab or in natural environment while working on one of the tasks
- No communication during observation

\+ Helps discover big problems

– No understanding of decision process (that may be wrong) or user's mental model, opinions, or feelings

E7: Think Aloud



Saul Greenberg

- As E6, but user is asked to say aloud
  - What he thinks is happening (state)
  - What he is trying to achieve (goals)
  - Why he is doing something specific (actions)
- Most common method in industry

\+ Good to get some insight into user's thinking, but:

  - Talking is hard while focusing on a task
  - Feels weird for most users to talk aloud
  - Conscious talking can change behavior

E8: Constructive Interaction

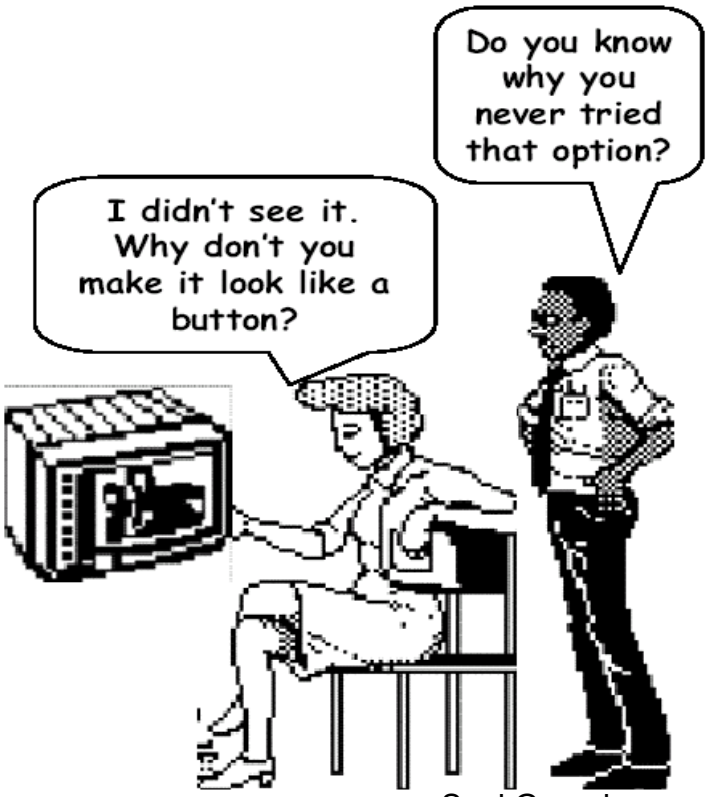
Saul Greenberg

- Two people work on a task together
  - Normal conversation is observed (and recorded)
  - More comfortable than Think Aloud
- Variant of this: Different partners
  - Semi-expert as “trainer”, newbie as “student”
  - Student uses UI and asks, trainer answers
  - Good: Gives insight into mental models of beginner and advanced users at the same time!

Recording Observations

- Paper and pencil
  - Evaluator notes events, interpretations, other observations
  - Cheap but hard with many details (writing is slow)
  - Forms can help
- Audio recording
  - Good for speech with Think Aloud and Constructive Interaction
  - But hard to connect to interface state
- Video
  - Ideal: 2 cameras (user and screen) in 1 picture
  - Best capture, but may be too intrusive initially
- Logging
  - Log input events of the user, synchronize with audio & video

E9: Retrospective Testing



- Additional activity after an observation
- Subject and evaluator look at video recordings together, user comments his actions retrospectively
- Good starting point for subsequent interview, looking at video avoids wrong memories
- Often results in concrete suggestions for improvement

E10: Controlled Experiments

- Quantitative, empirical method
- Steps
  - Formulate hypothesis
  - Design experiment, pick variable and fixed parameters
  - Choose subjects
  - Run experiment
  - Interpret results to accept or reject hypothesis

E10: Controlled Experiments

- Subjects
  - Similar to real users in profile
    - Age, education, computer and domain expertise, system knowledge,…
  - Use at least 10 subjects
  - Use more if you need finer details
- Variables
  - Independent: are varied under your control
    - E.g., font size
  - Dependent: are measured
    - E.g., execution time, error rates, subjective preferences

Hypothesis

- A claim that predicts outcome of experiment
  - Example: Reading text in capital letters takes longer than in reading text in small letters
- Hypothesis claims that changing independent variables influences dependent variables
  - Example: Changing small to capital letters (independent variable) influences reading time (dependent variable)
- Experimental goal: Confirm hypothesis
- Approach: Reject null hypothesis (inverse, i.e., “no influence”)
  - Null hypothesis is a term from statistical testing: The samples are drawn from the same statistical distribution

Choosing a Method

- Between-groups
  - – Each subject only does one variant of the experiment
  - – There are at least 2 variants
    (manipulated form & control, to isolate effect of manipulation)
  - \+ No learning effect across variants
  - – But requires more users
- Within-groups
  - – Each subject does all variants of the experiment
  - \+ Less users required, individual differences canceled out
  - – But often learning effect across variants problem

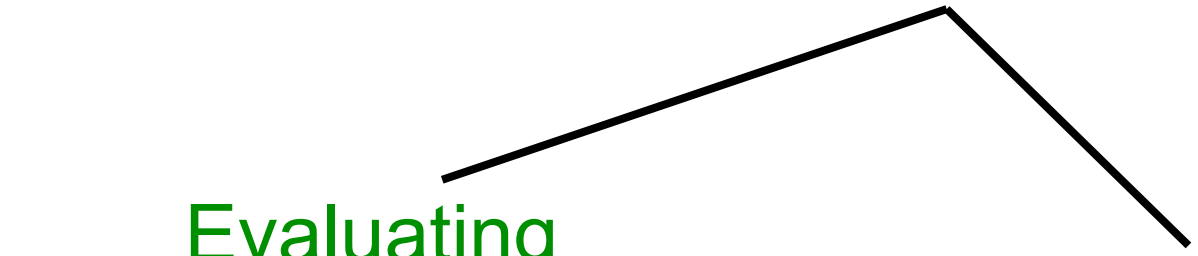
Analyzing Results

- Statistical analysis
  - Often assumptions about underlying distribution
  - t-test: Compare two groups, normal distribution
  - Analysis of variance (ANOVA): Compare two or more groups, normal distribution
  - Regression analysis: How well does result fit to a model?
  - Wilcoxon- or Mann/Whitney test, X^2 test
- Choice depends on
  - Number, continuity, and assumed distribution of dependent variables
  - Desired form of the result (yes/no, size of difference, confidence of estimate)

Other Evaluation Methods

- Before and during the design, with users
  - Personal interviews
  - Questionnaires
- After completing a project
  - Email bug report forms
  - Hotlines
  - Retrospective interviews and questionnaires
  - Field observations (observe running system in real use)

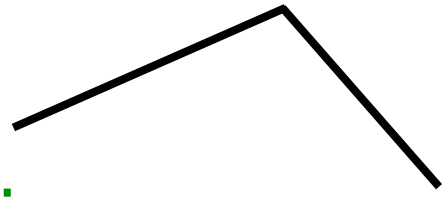
Evaluation Techniques



Evaluating Without Users

E1 Literature Review
E2 Cognitive Walkthrough
E3 Heuristic Evaluation
E4 Model-Based Evaluation

Evaluating With Users



Qualitative

E5 Conceptual Model Extraction
E6 Silent Observation
E7 Think Aloud
E8 Constructive Interaction
E9 Retrospective Testing

Quantitative

E10 Controlled Experiments

+ Interviews, questionnaires,...

Interviews

- Unstructured
  - Not directed by a script
  - Rich but not replicable
- Structured
  - Tightly scripted, often like a questionnaire
  - Replicable but may lack richness
- Semi-structured
  - Guided by a script but free to explore interesting issues in more depth
  - Good balance between richness and replicability

How to Ask Questions

- Clear and simple, not too broad
  - "How do you like the UI?" is too general!
- Affording logical, quantitative answers
  - Bad questions give unusable or wrong answers
  - Open vs. closed questions
- Users don't always answer truthfully
  - Lack of knowledge, bad estimates, embarrassment
  - So formulate questions carefully, maybe indirectly
- No leading questions!
  - For initial input, do not focus on presenting your design ideas, but on learning about the task

Running the Interview

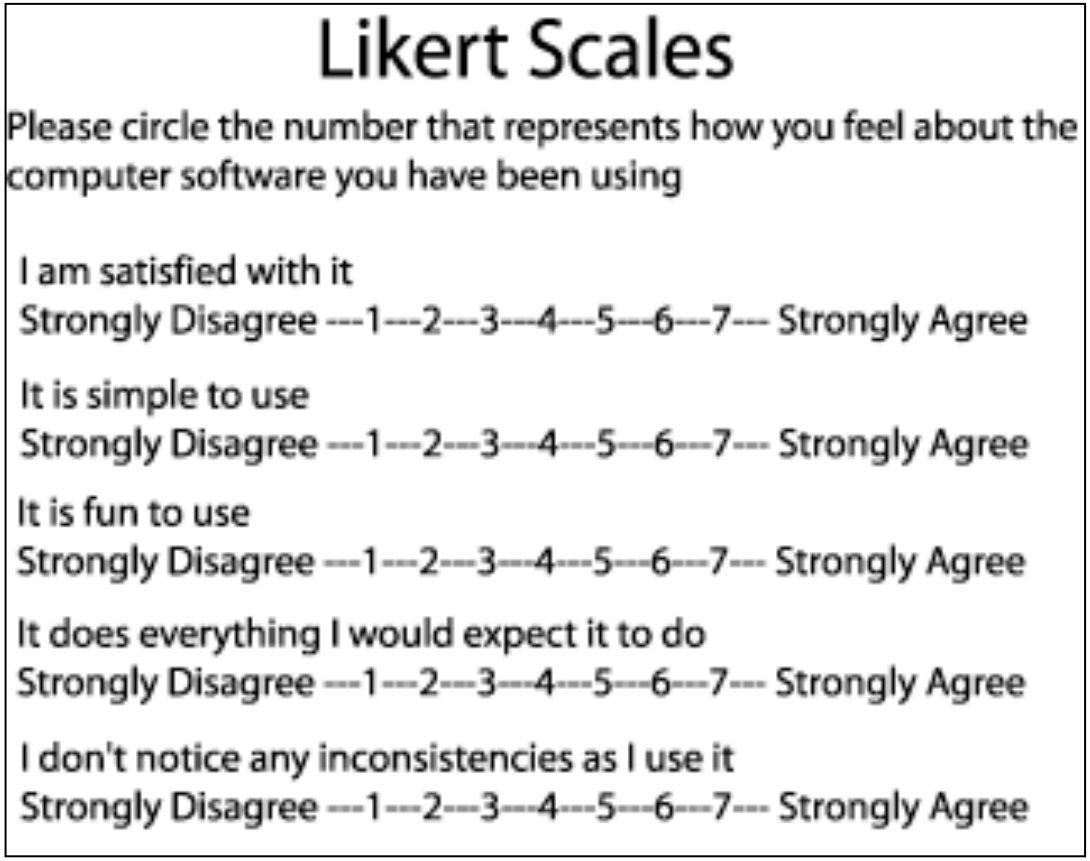
- Introduction
  - Introduce yourself, explain the goals of the interview, reassure about the ethical issues, ask to record, present any informed consent form
- Warm-up
  - Make first questions easy and non-threatening
- Main body
  - Present questions in a logical order
- A cool-off period
  - Include a few easy questions to defuse tension at the end
- Closure
  - Thank interviewee, signal the end, e.g., switch recorder off

Questionnaires

- Can be administered to large populations
  - Paper, email, and the web used for dissemination
- Provide clear instructions on how to complete the questionnaire
- Decide on whether phrases will all be positive, all negative, or mixed
- Presentation consistency:
  - Yes/No always in same position
    - yes ( ) no ( )
  - All positives always at same side
    - Bad |--|--|--|--|--|--| Good
  - Avoid conflict with existing scoring systems
    - 1-6 or A-F for grades

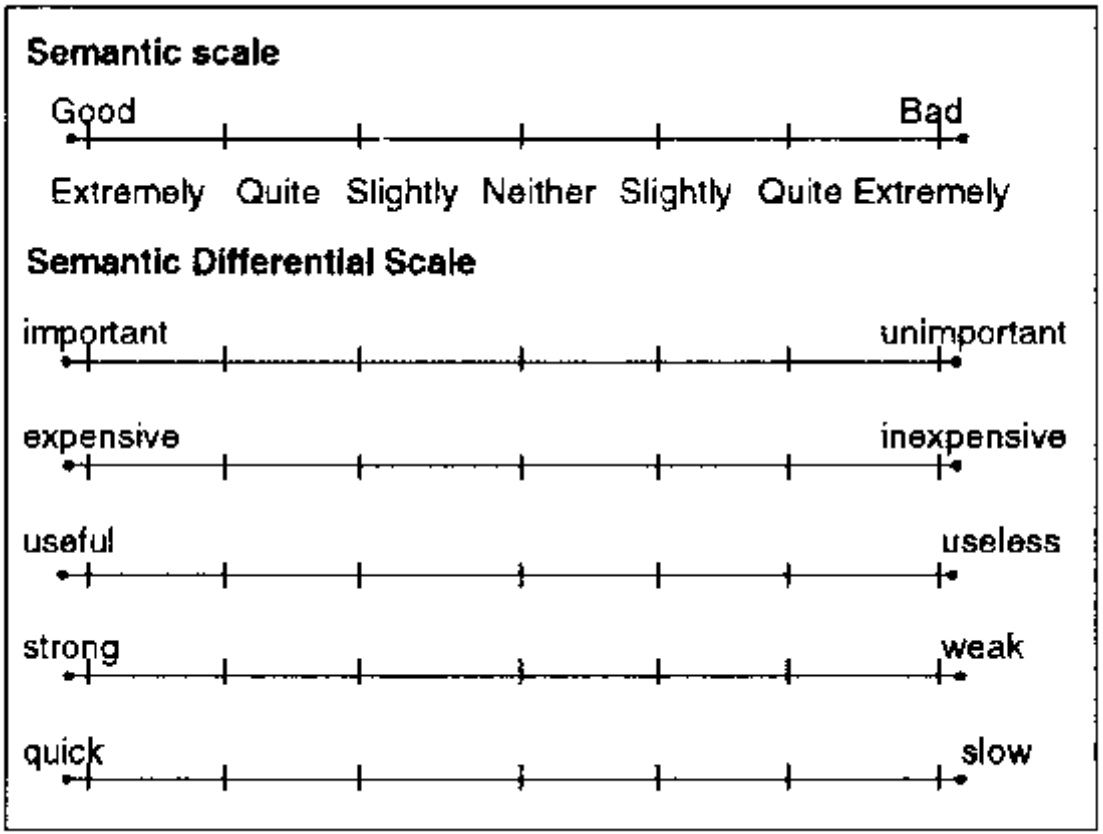
Likert Scales

- Measures degree of agreement with a statement
- Widely used for measuring opinions, attitudes, beliefs



Semantic Differential Scales

- Range of bipolar attitudes about a particular item
- Pair of attitudes represented as pair of adjectives



Combining Techniques

- Example: Combining interviews and questionnaires
  - Interviews with core group of users
  - Questionnaires for wider group of stakeholders
  - Interviews face-to-face
  - Questionnaires via email

- Triangulation: Use different approaches and perspectives to understand a problem or situation

Evaluation in the Mobile Context

- Context of use needs to be taken into account
  - Factors: User, activity, device, environment
- Usage “on the move”
  - Physically moving: walking, driving a car, traveling as a passenger
  - Being in different places: away from office environment or home
- Difficult to collect data in the field
  - Recording data
  - Controlling experimental conditions
- Dual-task interaction
  - E.g.: text input while walking
  - Test users reported less problems while walking (Kjeldskov et al.)

Example: Evaluating Attentional Resources in Mobile HCI

- Evaluating the competition for cognitive resources when mobile
- Field study in urban environment
  - Performance of mobile Web tasks
  - Movement through urban situations
- Attention during loading a page
  - Duration of continuous attention
    - Lab: 16.2s → field: 4s
  - Number of attention switches
    - Lab: 1 → field: 8
  - Switching-back durations
    - Railway station: 7-8s, quiet street: 4-6s

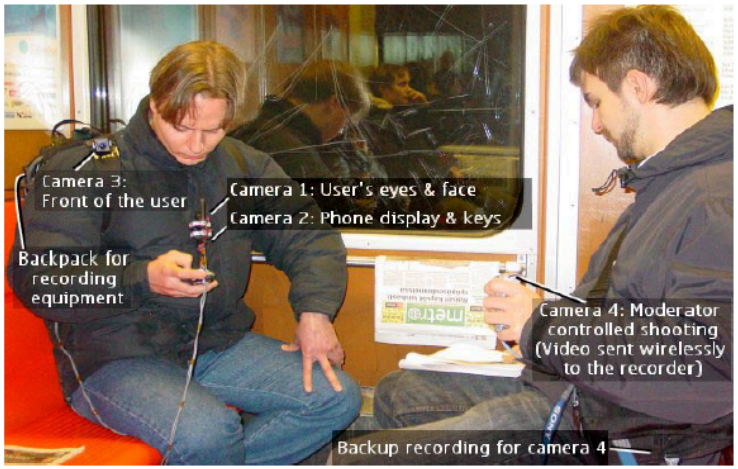


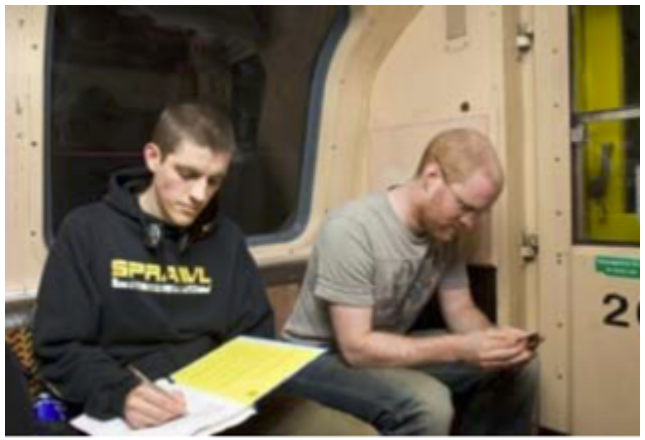
Figure 2. Configuration of recording equipment.

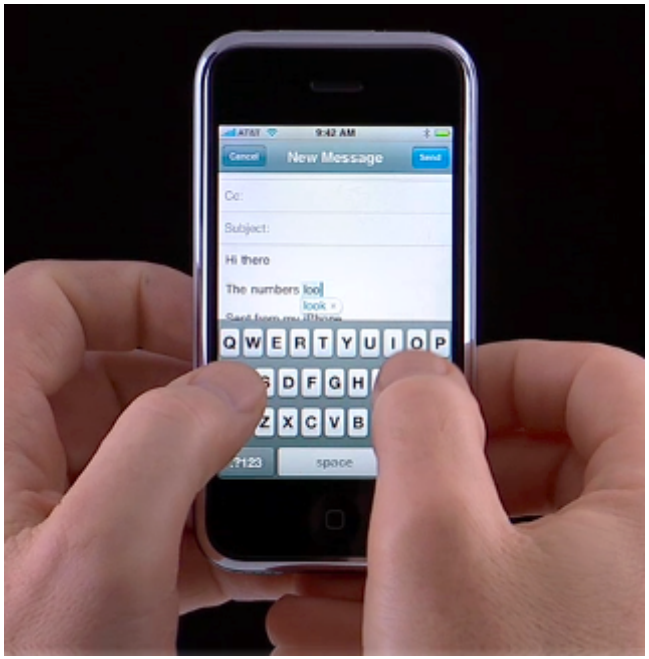
Figure 3. Output video data integrated on-the-fly.

Oulasvirta, Tamminen, Roto, Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. CHI '05.

Example: Text Input While on the Train

- Touchscreen phones have no tactile feedback for buttons
  - More errors typing text and numbers
- Performance comparison of physical buttons, touchscreen, and touchscreen+tactile
  - In lab and on subway
- Touchscreen+tactile as good as physical buttons
  - Touchscreen alone was poor





Brewster, Chohan, Brown: Tactile feedback for mobile interactions. CHI '07.

Example: Evaluation while Walking

- Evaluating preferred tapping time relative to gait phase
- Crossan et al.: Gait Phase Effects in Mobile Interaction, CHI '05

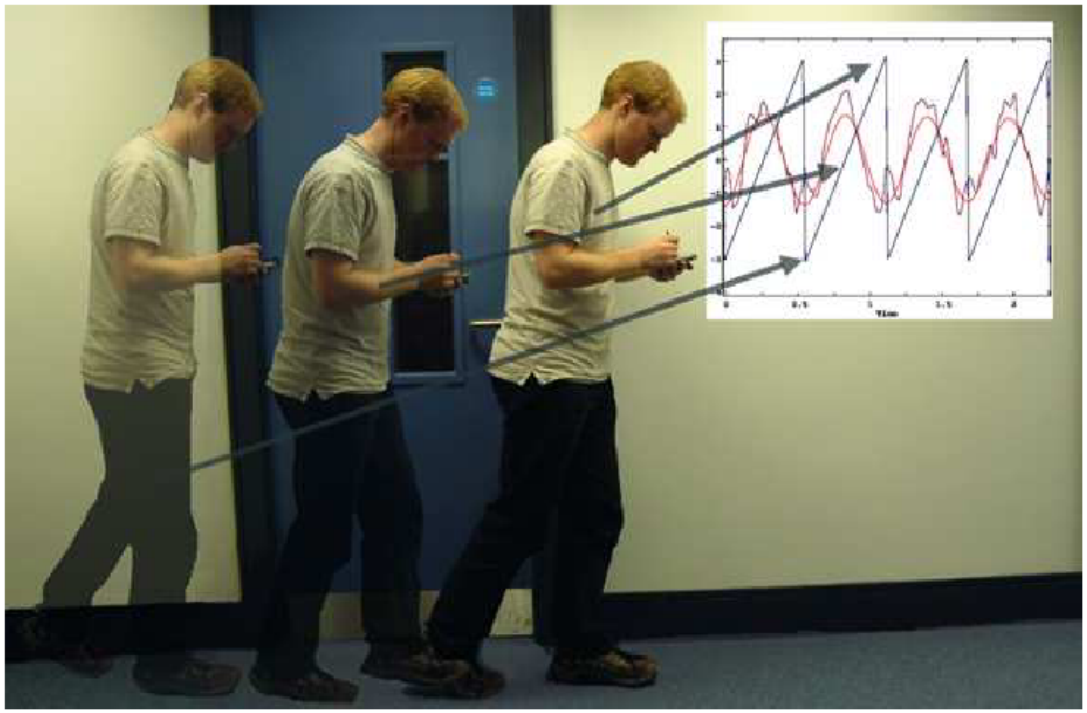


Figure 2. A user walking with the device and corresponding acceleration trace. The unfiltered vertical acceleration signal (rough sinusoid), the filtered signal (smooth sinusoid) and the phase estimate (in radians) for the signal (saw-tooth).

Crossan et al.: Gait Phase Effects in Mobile Interaction, CHI '05

Recording Video in Mobile Evaluation

- Noldus mobile device camera (right)
  - Wireless
- Google setup (left)
  - Observes display and keypad
- Useful if no access to application source code

Schusteritsch, Wei, LaRosa: Towards the perfect infrastructure for usability testing on mobile devices. CHI '07.

www.noldus.com

Summary

- Evaluate to ensure system matches users' needs
- Evaluation should happen throughout the design process
  - By design team (analytically)
  - By users (experimentally)
- A plethora of methods to evaluate designs
  - Decide when to apply which
- Treat testers with respect at all times!

Recommended Reading

- For the assignment:

- Mobile Web 2009 = Desktop Web 1998

  “Mobile phone users struggle mightily to use websites, even on high-end devices. To solve the problems, websites should provide special mobile versions.”

- http://www.useit.com/alertbox/mobile-2009.html