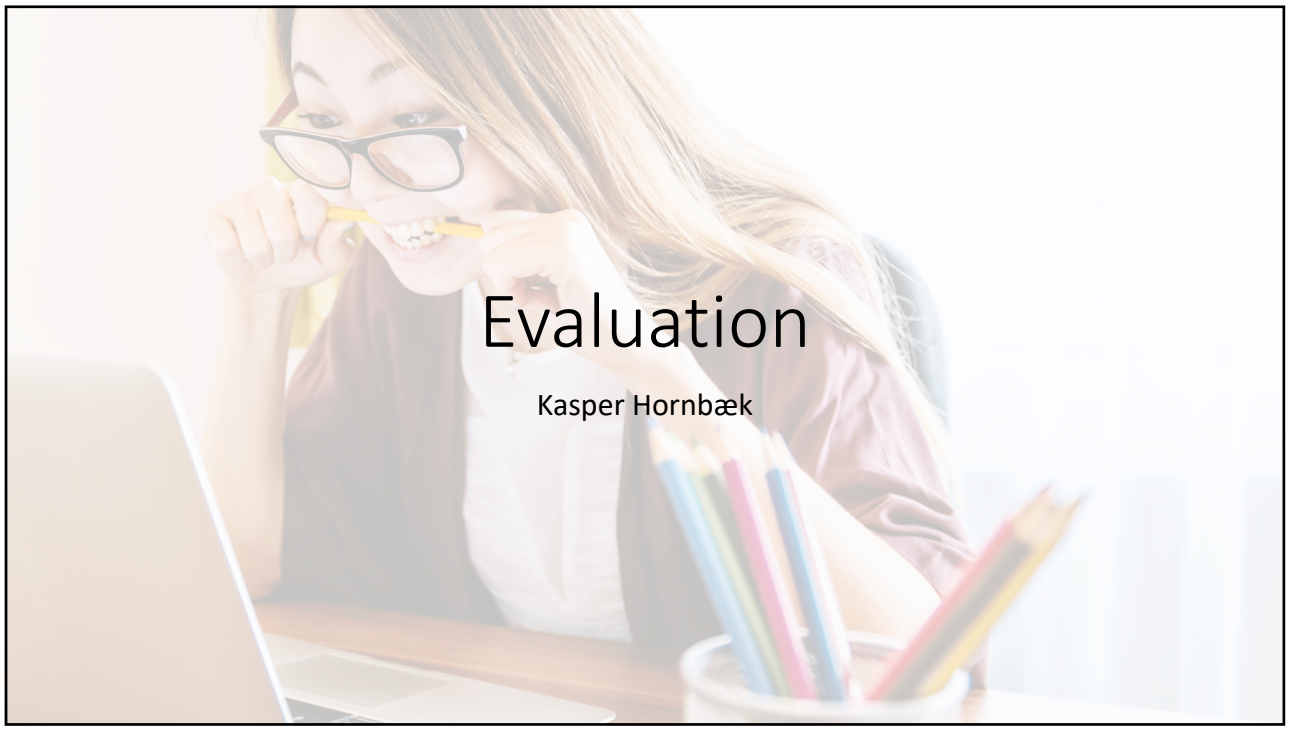


Evaluation

Kasper Hornbæk

1

Praktisk

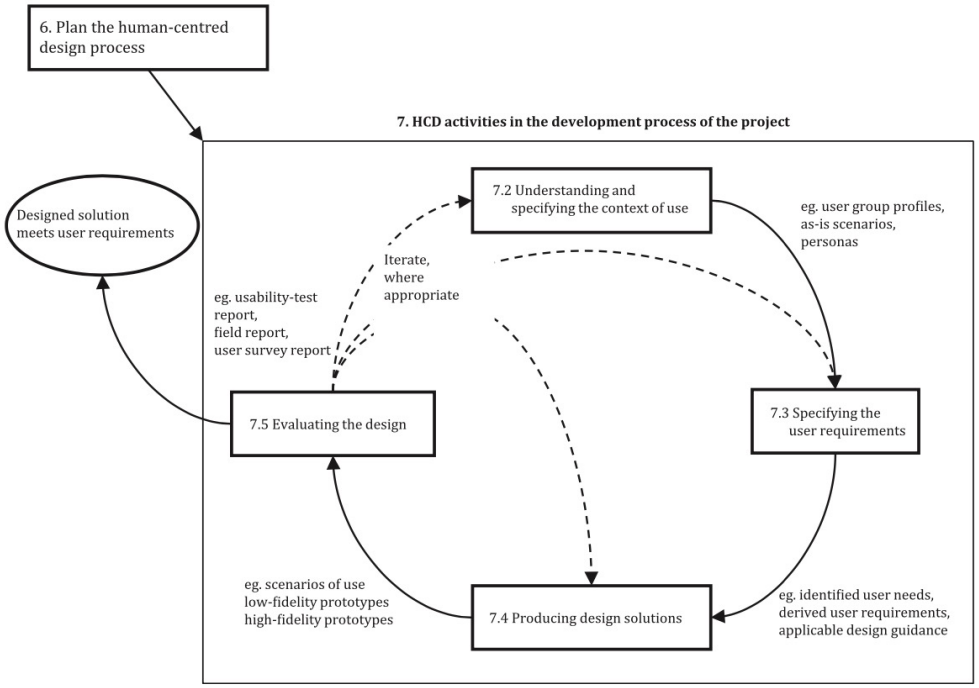
- Husk feedback på https://forms.gle/wr4YfhnSen8UQZ9CA

2

Det får I med i dag

- Hvorfor er det vigtigt at teste brugbarhed?
- Et overblik over beslutninger i evaluering
- Noget om tænke-højt studier

3



4

Why bother with testing?

- We check the correctness of computer programs. We should also check the usability of programs
- Impossible to make interactive systems correct in the first attempt

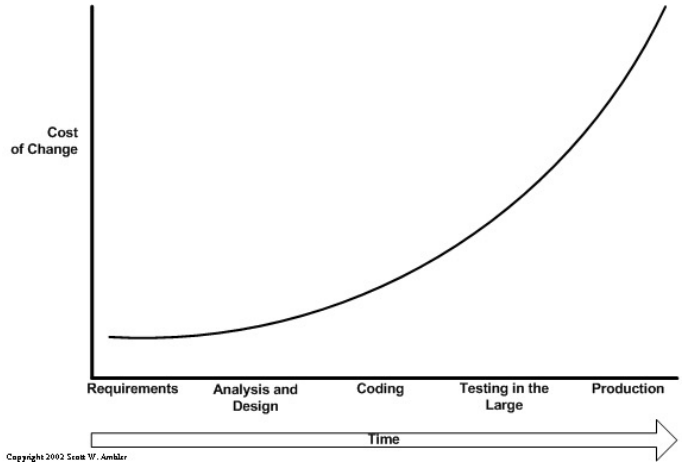
5

Why bother with testing?

- We check the correctness of computer programs. We should also check the usability of programs
- Impossible to make interactive systems correct in the first attempt
- Evaluation is worth it

6

Expensive to fix errors later



7

Cost-benefit argument

In the second example, the value of the usability engineering completed for Application 2, an internal product built to replace a manual process, is calculated. For Application 2, three iterative usability tests were conducted on a system that is employed by users to complete 1,308,000 tasks a year (Karat, 1990a). The usability work resulted in an average reduction of 9.6 minutes per task. The first-year benefits in increased productivity due to usability were as follows:

- Reduction in time on task from first to final user interface design for Application 2:

  Average savings = 9.6 minutes

- Projected first-year benefits due to increased user productivity:

  Tasks per year = 1,308,000

  1,308,000 x 9.6 minutes = 209,208 hours

  209,280 hours x Productivity ratio x Personnel costs = $6,800,000

The usability work for Application 2 included a benchmark field test, development of a high-fidelity prototype, and three iterations of usability design and testing of the user interface for the application. The usability work was completed across 21 months.

- Usability resource = $23,300
- Participant travel = $ 9,750
- Test-related development work = $34,950
- Total cost for 21 months = $68,000

The cost-benefit ratio of usability work is $6,800,000/$68,000 or a cost-benefit ratio of 1:100 for the first year of application deployment.

Karat (1998), p. 774

8

Why bother with testing?

- We check the correctness of computer programs. We should also check the usability of programs
- Impossible to make interactive systems correct in the first attempt
- Evaluation is worth it
- Evaluation helps involve users

9

10

Today

- Why do we evaluate?
- What are the key questions in evaluation?
- Analytic evaluations
- Empirical evaluations
- Impact of evaluations

11

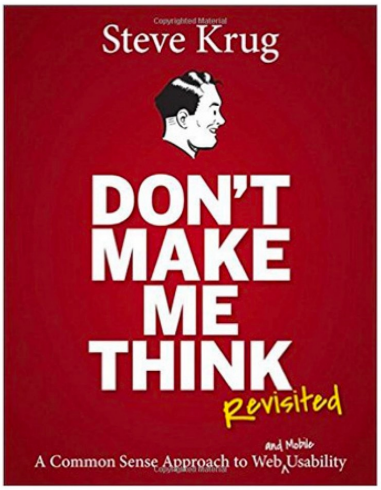
Who?

- Experts or users
  - Analytic versus empirical methods

12

Yardsticks of evaluation?

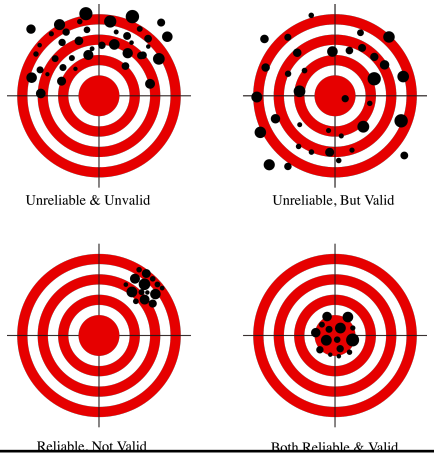
- Which criteria to use?
  - Usability measures
  - Goals (e.g., “being able to learn in 5 minutes”)
  - User experience measures
  - Guidelines for usable systems
  - Making people stop and think



13

When are evaluations good?

- Correct (aka valid)?
- Reliable?
- Realistic?
- Convincing?
- Useful to act on/diagnostic?



14

Purpose of evaluation?

- Summative versus formative evaluations

15

When?

- All the time!

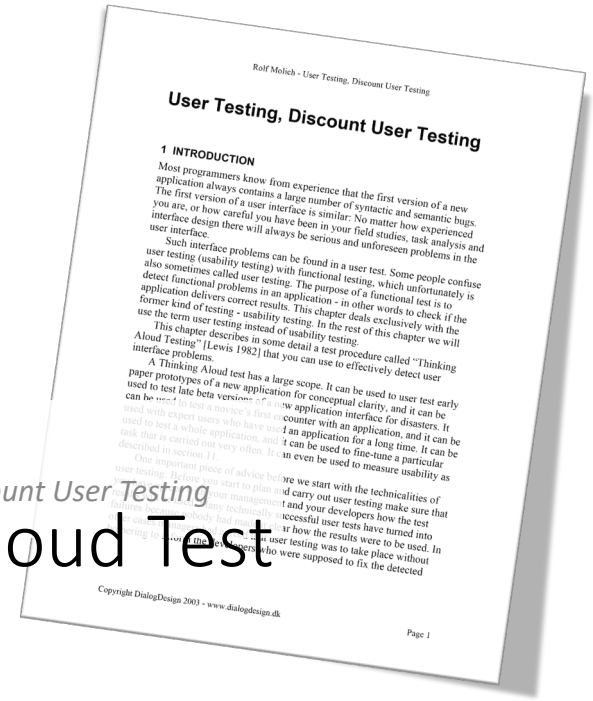
16

Choosing methods

Method	Definition	Pros	Cons
Heuristic evaluation	An analytic evaluation method in which evaluators go through an interface using a list of features of good user interfaces	Inexpensive	False positives
Think aloud testing	Users verbalize what they think about while they solve tasks with an interactive system; the thinking aloud is analyzed to find usability problems.	Inexpensive, convincing	Short-term
Usability test	An evaluation of the usability of an interactive system with representative users doing representative tasks	Direct assessment of usability	Misses the broader context of use
Experiment	An experimental comparison of the usability of at least two user interfaces	High precision; clean comparisons	Limited realism
Deployment study	Measure evaluation criteria after deployment with real user	Problems are real	Expensive

17

18



Rolf Molich - User Testing, Discount User Testing

User Testing, Discount User Testing

1 INTRODUCTION

Most programmers know from experience that the first version of a new application always contains a large number of syntactic and semantic bugs. The first version of a user interface is similar: No matter how experienced you are, or how careful you have been in your field studies, task analysis and interface design there will always be serious and unforeseen problems in the user interface.

Such interface problems can be found in a user test. Some people confuse user testing (usability testing) with functional testing, which unfortunately is also sometimes called user testing. The purpose of a functional test is to detect functional problems in an application - in other words to check if the application delivers correct results. This chapter deals exclusively with the former kind of testing - usability testing. In the rest of this chapter we will use the term user testing instead of usability testing.

This chapter describes in some detail a test procedure called "Thinking Aloud Testing" [Lewis 1982] that you can use to effectively detect user interface problems.

A Thinking Aloud test has a large scope. It can be used to user test early paper prototypes of a new application for conceptual clarity, and it can be used to test late beta versions of a new application interface for disasters. It can be used to test a whole application, and it can be used to fine-tune a particular task that is carried out very often. It can even be used to measure usability as described in section 11.

One important piece of advice before we start with the technicalities of user testing: Before you start to plan and carry out user testing make sure that your management and your developers how the test results can be used. Many technically successful user tests have turned into nothing because nobody had made up their mind how the results were to be used. In other words, the reason for doing user testing was to take place without the developers who were supposed to fix the detected

Copyright DialogDesign 2003 - www.dialogdesign.dk

Page 1

Rolf Molich: *User Testing, Discount User Testing*

Think-Aloud Test

19

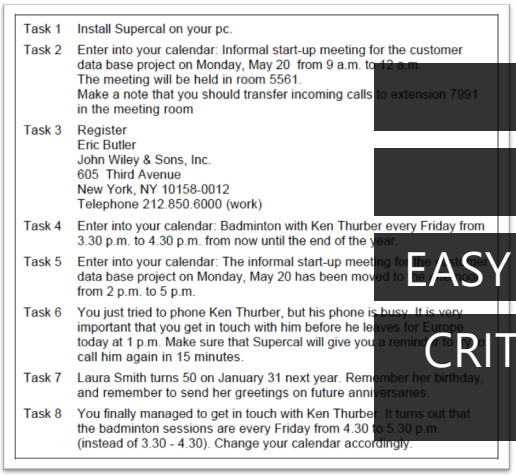
Think-Aloud Test

- **Verbalization**
  users verbalize their thoughts while performing a task
- **Capture**
  verbalized thoughts can be captured through:
  notes, audio, video, …

20

Planning think-aloud tests

1. Create scenarios and tasks



Task 1 Install Supercal on your pc.

Task 2 Enter into your calendar: Informal start-up meeting for the customer data base project on Monday, May 20 from 9 a.m. to 12 a.m. The meeting will be held in room 5561. Make a note that you should transfer incoming calls to extension 7991 in the meeting room

Task 3 Register
Eric Butler
John Wiley & Sons, Inc.
605 Third Avenue
New York, NY 10158-0012
Telephone 212.850.6000 (work)

Task 4 Enter into your calendar: Badminton with Ken Thurber every Friday from 3.30 p.m. to 4.30 p.m. from now until the end of the year.

Task 5 Enter into your calendar: The informal start-up meeting for the customer data base project on Monday, May 20 has been moved to [illegible] from 2 p.m. to 5 p.m.

Task 6 You just tried to phone Ken Thurber, but his phone is busy. It is very important that you get in touch with him before he leaves for Europe today at 1 p.m. Make sure that Supercal will give you a reminder to call him again in 15 minutes.

Task 7 Laura Smith turns 50 on January 31 next year. Remember her birthday, and remember to send her greetings on future anniversaries.

Task 8 You finally managed to get in touch with Ken Thurber. It turns out that the badminton sessions are every Friday from 4.30 to 5.30 p.m. (instead of 3.30 - 4.30). Change your calendar accordingly.

REALISTIC

NO CUES

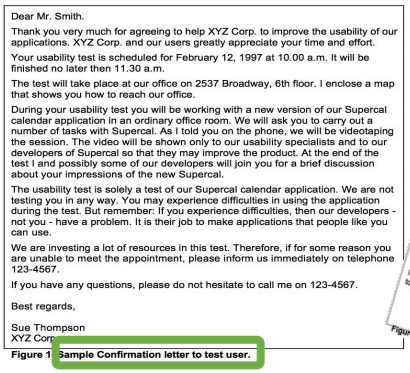
EASY FIRST TASK

CRITERION FOR DONE

21

Planning think-aloud tests

1. Create scenarios and tasks
2. Recruit and inform users



Dear Mr. Smith.

Thank you very much for agreeing to help XYZ Corp. to improve the usability of our applications. XYZ Corp. and our users greatly appreciate your time and effort.

Your usability test is scheduled for February 12, 1997 at 10.00 a.m. It will be finished no later then 11.30 a.m.

The test will take place at our office on 2537 Broadway, 6th floor. I enclose a map that shows you how to reach our office.

During your usability test you will be working with a new version of our Supercal calendar application in an ordinary office room. We will ask you to carry out a number of tasks with Supercal. As I told you on the phone, we will be videotaping the session. The video will be shown only to our usability specialists and to our developers of Supercal so that they may improve the product. At the end of the test I and possibly some of our developers will join you for a brief discussion about your impressions of the new Supercal.

The usability test is solely a test of our Supercal calendar application. We are not testing you in any way. You may experience difficulties in using the application during the test. But remember: If you experience difficulties, then our developers - not you - have a problem. It is their job to make applications that people like you can use.

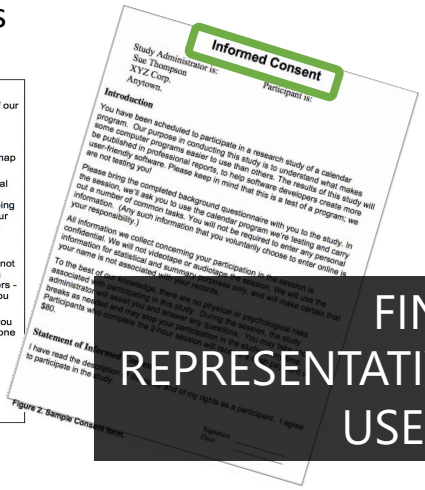
We are investing a lot of resources in this test. Therefore, if for some reason you are unable to meet the appointment, please inform us immediately on telephone 123-4567.

If you have any questions, please do not hesitate to call me on 123-4567.

Best regards,

Sue Thompson
XYZ Corp.

Figure 1. Sample Confirmation letter to test user.



Informed Consent

Study Administrator is:
Sue Thompson
XYZ Corp.
Anytown.

Participant is:

Introduction

You have been scheduled to participate in a research study of a calendar program. Our purpose in conducting this study is to understand what makes some computer programs easier to use than others. The results of this study will be published in professional reports, to help software developers create more user-friendly software. Please keep in mind that this is a test of a program; we are not testing you!

Please bring the completed background questionnaire with you to the study. In the session, we'll ask you to use the calendar program we're testing and carry out a number of common tasks. You will not be required to enter any personal information. (Any such information that you voluntarily choose to enter online is your responsibility.)

All information we collect concerning your participation in the session is confidential. We will not videotape or audiotape the session. We will use the information for statistical and summary purposes only, and will make certain that your name is not associated with any records.

To the best of our knowledge, there are no physical or psychological risks associated with participating in this study. During the session, the study administrator may stop your participation in the study. You may also take breaks as needed and may skip your participation in the. You may quit Participants will receive $80 if they complete the 2-hour session in the. You may quit.

Statement of Informed Consent

I have read the description of the study and of my rights as a participant. I agree to participate in the study.

Signature
Date

Figure 2. Sample Consent form.

FIND REPRESENTATIVE USERS

22

Planning think-aloud tests

1. Create scenarios and tasks
2. Recruit and inform users
3. Select location
4. Practice the test

DON'T WASTE PEOPLES' TIME

Source: MavInnovationLab

23

Conducting think-aloud tests

1. Create a good atmosphere

24

Conducting think-aloud tests

1. Create a good atmosphere
2. Decide the degree of verbalization
   - Pure think-aloud (Ericsson & Simon)
   - Molich's advice
     - What do you want to do?
     - What were you expecting to happen?
     - What is the system telling you?
     - Why has the system done that?
     - What are you doing now?

THIS IS NOT AN INTERVIEW

25

Conducting think-aloud tests

1. Create a good atmosphere
2. Decide the degree of interaction
3. Document the test

USE SCREEN RECORDER

POST-TEST INTERVIEW

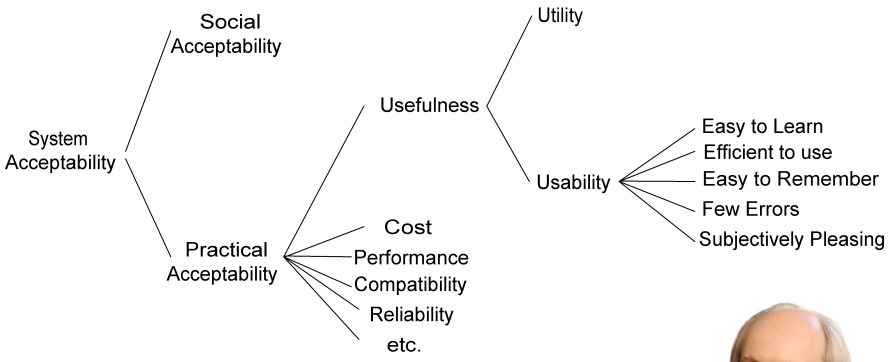
WRITE UP NOTES IMMEDIATELY

26

What are we looking for?

27

Usability: Nielsen 1993



JAKOB NIELSEN
NN group

28

What is a usability problem?

Box 25.1 **Characteristics of Usability Defects**

A usability defect has one or more of the following characteristics:

- It irritates or confuses the user.
- It makes a system hard to install, learn, or use.
- It causes mental overload of the user. (For example, the user may have to think a lot as the required action or feedback from the system may not be obvious or sufficient.)
- It causes poor user performance.
- It violates design standards or guidelines.
- It reduces trust or credibility of the system.
- It tends to cause repeated errors.
- It could make the system hard to market.

Stone et al. (1995), box 25.1

29

What is a usability problem?

(1) The user articulates a goal and cannot succeed in attaining it within three minutes,

(2) the user explicitly gives up,

(3) the user articulates a goal and has to try three or more actions to find a solution,

(4) the user produces a result different from the task given,

(5) the user expresses surprise,

(6) the user expresses some negative affect or says something is a problem,

(7) the user makes a design suggestion

Jacobsen et al. (1998)

30

Analysis of tests

- Analysis and interpretation of tests often missing in practice (Nørgaard & Hornbæk 2006)
  - Be explicit about usability problems
  - Discuss the causes of problems
  - Sketch one or more solutions to the problems

31

Problem description

- Analysis of a Problem
  frequency: is the problem common or rare?
  impact: can users overcome in an easy way or is it difficult?
  persistence: will the problem repeatedly bother users or is it a one-time problem they can overcome once they know about it?

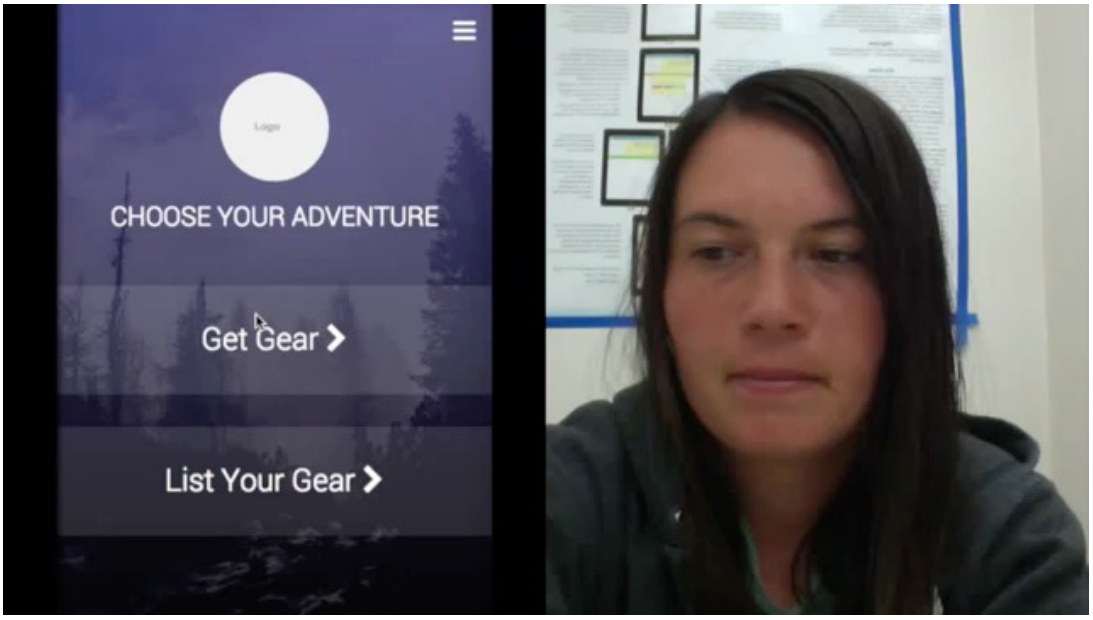
- Severity of a Problem
  1. Critical problem: Causes frequent disasters (see section 11.1). It is strongly recommended that this problem should be fixed before the application is released.
  2. Serious problem. Causes some disasters. Should be fixed in the next release.
  3. Cosmetic problem. Should be corrected when time permits.

32

Problem description, cont'd

- Please also explain how to fix the problem and be specific about the assumed cause of the problem

33



34

Think-Aloud tests: Pros

- Cheap
- Persuasive: developers and stakeholders believe them
- Valid: problems found are mostly real problems

35

Think-Aloud Tests: Cons

- Influences on the task
  it may change how participants will solve the problem
- Artificial situation
  may appear unnatural and/or uncomfortable for participants
- Cognitive load
  difficult to do, e.g., if the task is difficult (participants concentrate)
- Limited Possibilities
  focuses on short-term interactions

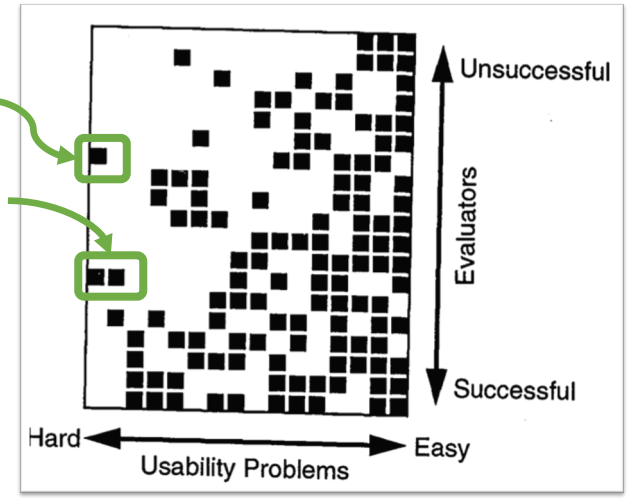
36

Cons: Variability among evaluators

The hardest-to-find usability problems are found by evaluators who do not otherwise find many usability problems

Issues with reliability

Involve multiple evaluators



37

Empirical evaluation methods

- Involve users
- Collect data about their performance and opinions
- Analyze to find problems and remedies

- Include
  - Think-aloud test (Molich, 2003)
  - Cooperative evaluation
  - Controlled experiments
  - Living labs and in-situ evaluations
  - Usability test

SUMMATIVE OR FORMATIVE

38

39

Business goals in usability evaluation

- 44 novice evaluators were asked to conduct think aloud tests (Molich 2007)
- All evaluators received material describing business goals
- Evaluated an e-commerce site

Hornbæk, K. & Frøkjær, E. (2008), *CHI'08*

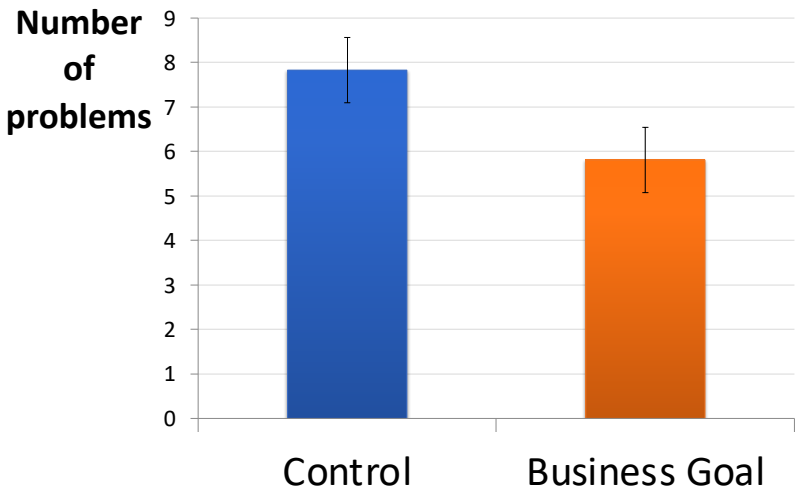
40

Experiment, cont.

- Half of the evaluators were asked to consider business goals
  - During planning
  - During reporting
- The other half received no special instructions (control)

41

Problem counts



Hornbæk, K. & Frøkjær, E. (2008), *CHI'08*

42

IBM's assessment

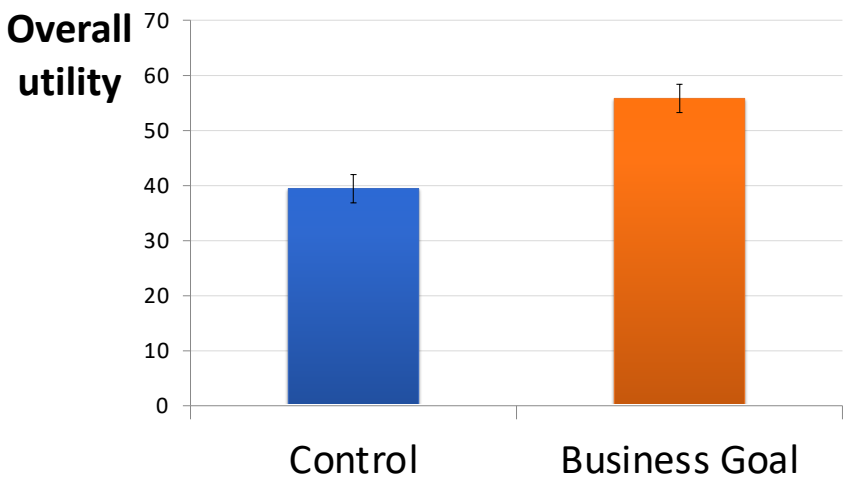
- Company assessed 290 problems identified on four dimensions of downstream utility

Clarity	incomprehensible	crystal clear
Interaction	irrelevant	very important
Business	irrelevant	very valuable
Overall usefulness	irrelevant	very useful

Hornbæk, K. & Frøkjær, E. (2008), *CHI'08*

43

IBM's assessment, con't



Hornbæk, K. & Frøkjær, E. (2008), *CHI'08*

44

Way to impact:

- Problem counting would hide the main finding
- Downstream utility seems more interesting than following the problem counting dogma
- Context of evaluation matters a lot: small difference with a large impact.

45

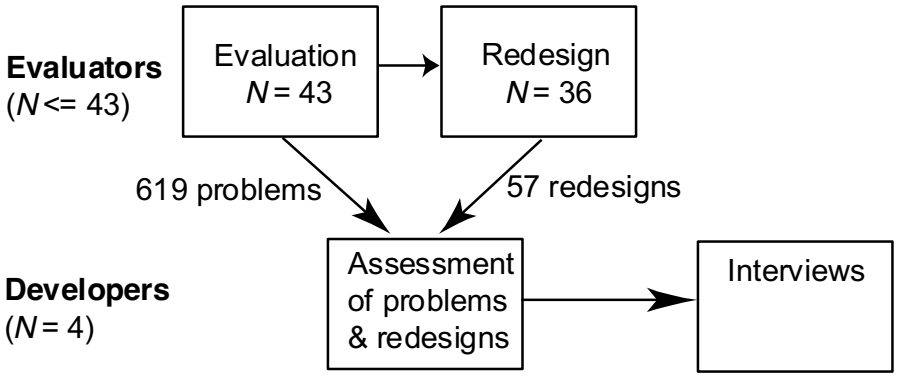
Limitations of focusing on problems

- Defect-identification typical for Westerners (deBono 1994)

  "This is a basic habit of Western thinking. If you can only get rid of faults everything else will be fine. There are two obvious dangers: 1. We only focus our thinking on what is wrong. 2. Getting rid of the faults in a poorly designed system does not result in a better-designed system."
- Low practical significance (Wixon 2003)
- Poor way of communicating?

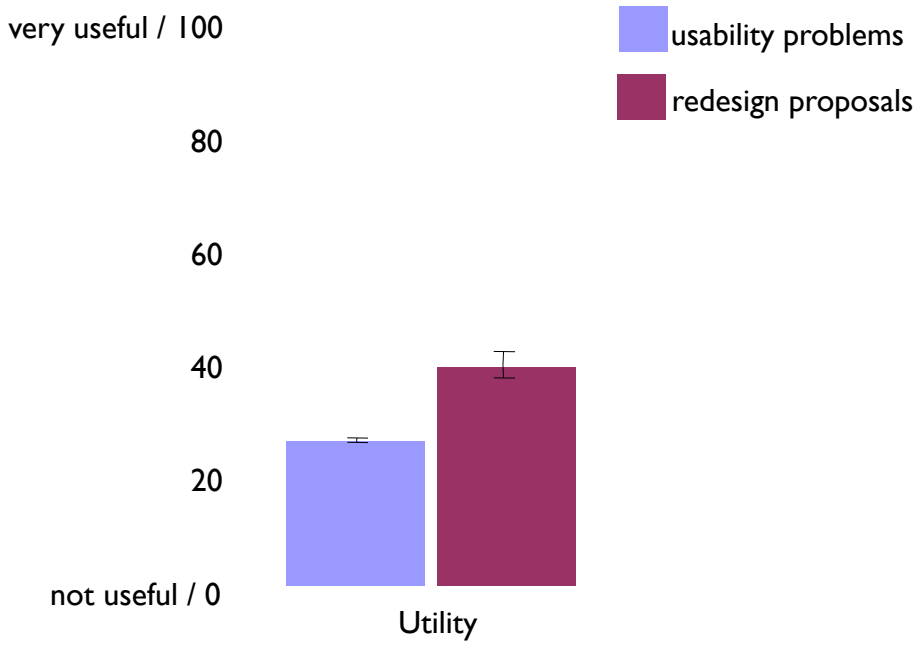
46

What about focusing on redesigns?



Hornbæk, K. & Frøkjær, E. (2005), *CHI 2005*, 391-400

47



Hornbæk, K. & Frøkjær, E. (2005), *CHI 2005*, 391-400

48

Interviews: redesign proposals

- Redesigns give ideas
  - "ok, there were some pearls in it … sometimes things that we had not thought about, especially redesign proposals for saying, ok that way of doing it is also possible"
- More concrete than usability problems
- More constructive than usability problems
  - "it is almost obvious that it is better to say: if it were this way it was better, rather than just saying: this is wrong"

Hornbæk, K. & Frøkjær, E. (2005), *CHI 2005*, 391-400

49

What Do Usability Evaluators Do in Practice?

- Think aloud testing is the most widely used evaluation method
- Almost no studies of how professional evaluators actually conduct and report think aloud tests
  - but see Boren & Ramey 2000 and Molich et al. 2004
- Would help researchers understand practice and aid practitioners in improving their work habits
  - Are we supporting the right activities? Are techniques misunderstood or used as prescribed?

50

What do evaluators really do?

- 14 think aloud sessions in seven companies:
  - 2-8500 employees
  - 2-8 of the employees work with evaluation
  - Evaluators' experience between 1 and 8 years
- We followed the setup, carrying out, analysis and discussion of the test
- About 25 hours of recordings

Nørgaard, M. & Hornbæk, K. (2006), "What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing", *ACM Symposium on Designing Interactive Systems (DIS 2006),* 209-218.

52

Main findings

Area of attention	*Main finding*	*N*	*Example of observations and quotes*
Analysis of results from a test session	Analysis is unstructured	9	Scattered fragments of analysis; no systematic approach used
	Analysis is incomplete	9	Does not identify causes or solutions; restricts discussion to user traits
	Analysis as a summary with the user	3	"Let's sum up"; selecting a few problems for further questioning; listing key findings
Confirmation of known issues as a test' focus	Looking for known issues	8	"Now, I am just looking for ammunition"; develops ideas of problems before testing; tasks and questions designed to point out known issues
	Practitioners have foreseen problems	5	"We have a gut feeling", "I told you so"
Practical realities influencing tests	Technical problems	8	System breaking down; long response times in test environment; installation or security messages interrupt workflow
	Unfinished prototypes	6	Parts of prototype missing or inaccessible; "a log-in name should not be WaddleFish"; texts and pictures are wrong or out of date
Questions asked during a test	Problems are explained, not experienced	13	"Do you think you would go back to the front page"; "did you notice this column"; "what do you expect to see"
	Leading questions	13	Questions address certain parts of GUI or system; evaluator hints the solution; "Can you do this in another way?"
	Unnecessary or obvious questions	10	"You did figure out to press the print button?"; asking user to locate information that clearly appears on the present screen; asking if user would like relevant information
Trying to meet laboratory-style scientific standards	Evaluators want similar conditions for users under test	5	"We have to make sure all users get the same questions"
	Rigid or artificial procedures	3	Laboratory style procedures; Danish evaluators speaking English to a Danish user; measuring subjective satisfaction overly systematic
Uncovering usability problems or utility concerns	User points to utility or lack thereof	10	"I would not do it like this"; user chooses to solve task without help of system
	Evaluator probes utility concerns	7	Asking about normal workflow; asking whether a task is realistic; "What would you typically do?"

53

How are think-aloud sessions analyzed?

- No sessions were immediately followed by an attempt at a structured analysis of the session
- Analysis seems a weak part of the think aloud setup
  - Challenge for researchers
  - Practitioners should more systematically collect and discuss their observations

54

Questions during the sessions

- Some questions are not used to understand usability problems that are experienced by the user, but ask the user to predict problems
- For these problematic questions:
  - Researchers must note that practitioners want to get a variety of information: can this be done in a valid way?
  - Practitioners should be cautious with respect to how they put questions to users

55

Utility versus usability

- We may distinguish usability and utility
- Utility concerns are less frequently reported compared to usability problems
- Utility problems seem very important:
  - Researchers should develop techniques to investigate utility
  - Practitioners could ask more about the utility of a system and about how users normally would solve a task

56

Main findings

Area of attention	*Main finding*	*N*	*Example of observations and quotes*
Analysis of results from a test session	Analysis is unstructured	9	Scattered fragments of analysis; no systematic approach used
	Analysis is incomplete	9	Does not identify causes or solutions; restricts discussion to user traits
	Analysis as a summary with the user	3	"Let's sum up"; selecting a few problems for further questioning; listing key findings
Confirmation of known issues as a test' focus	Looking for known issues	8	"Now, I am just looking for ammunition"; develops ideas of problems before testing; tasks and questions designed to point out known issues
	Practitioners have foreseen problems	5	"We have a gut feeling", "I told you so"
Practical realities influencing tests	Technical problems	8	System breaking down; long response times in test environment; installation or security messages interrupt workflow
	Unfinished prototypes	6	Parts of prototype missing or inaccessible; "a log-in name should not be WaddleFish"; texts and pictures are wrong or out of date
Questions asked during a test	Problems are explained, not experienced	13	"Do you think you would go back to the front page"; "did you notice this column"; "what do you expect to see"
	Leading questions	13	Questions address certain parts of GUI or system; evaluator hints the solution; "Can you do this in another way?"
	Unnecessary or obvious questions	10	"You did figure out to press the print button?"; asking user to locate information that clearly appears on the present screen; asking if user would like relevant information
Trying to meet laboratory-style scientific standards	Evaluators want similar conditions for users under test	5	"We have to make sure all users get the same questions"
	Rigid or artificial procedures	3	Laboratory style procedures; Danish evaluators speaking English to a Danish user; measuring subjective satisfaction overly systematic
Uncovering usability problems or utility concerns	User points to utility or lack thereof	10	"I would not do it like this"; user chooses to solve task without help of system
	Evaluator probes utility concerns	7	Asking about normal workflow; asking whether a task is realistic; "What would you typically do?"

57

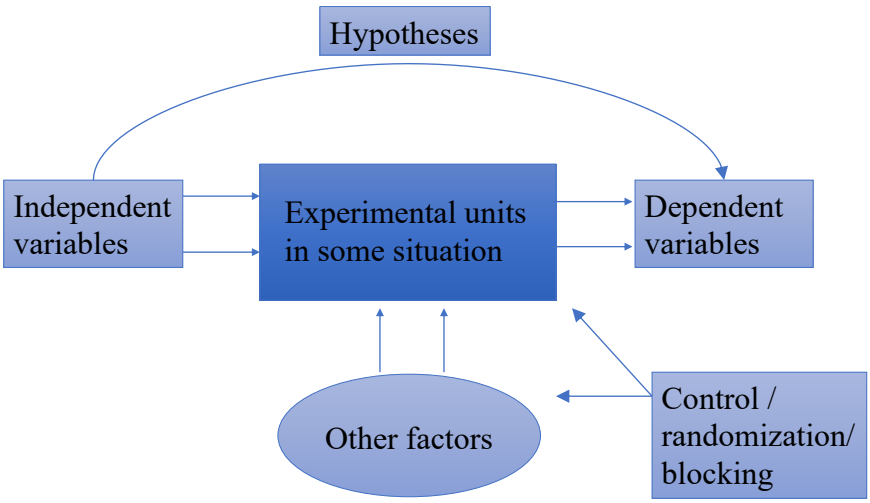
58

What is an experiment?

- "A test under controlled conditions that is made to demonstrate a known truth, examine the validity of a hypothesis, or determine the efficacy of something previously untried", *American Heritage Dictionary*
- Broad use, but usually understood stricter in research methodology

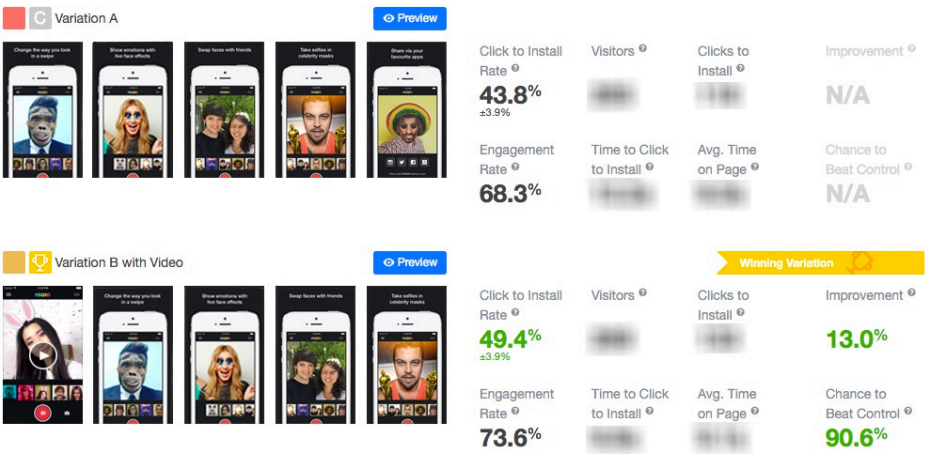
59

Schematics of experiments



60

A/B experiments



61

A/B example

- When you do a Google Search, you have these ads with the links that lead to other websites. Does the colour of the links matter as long as we are getting what we are searching for?
- Google showed 1% of the users a particular blue, another 1% a different blue and so on. Over 40 experiments were ran.
- Look at the click-rate for adds.
- The implication: an additional 200 million dollars in revenue due to this colour change

https://bbharathkumar.medium.com/data-driven-decisions-googles-50-shades-of-blue-experiment-996f01819a97

62

Randomised, controlled trials

- Experiments are well-known in medicine

63

How?

64

Formulating hypotheses

- Good hypotheses:
  - Are developed early
  - Define independent and dependent variables
  - May be motivated by theory or literature
  - Are concrete, specific, clear
  - Can be shown wrong
- Hypotheses help think about about one's work

65

Selecting independent variables

- Often use control group or baseline technique
- Are they:
  - Similar in all non-essential aspects?
  - Biased? Best possible alternative?
  - Representative?
  - Realistic?
- Use several alternatives

66

Selecting dependent variables

- Lots of measures in the literature, use them!
  - Usability (e.g., questionnaires)
  - Psychophysics (e.g., ranking procedures)
  - Graphics/image processing outlets (e.g., *ACM Transactions on Graphics*)
- Objective versus subjective measures

67

Dependent variables

Usability objective	*Effectiveness measures*	*Efficiency measures*	*Satisfaction measures*
Overall usability	Percentage of tasks successfully completed Percentage of users successfully completing tasks	Time to complete a task Time spent on non-productive actions	Rating scale for satisfaction Frequency of use if this is voluntary (after system is implemented)
Meets needs of trained or experienced users	Percentage of advanced tasks completed Percentage of relevant functions used	Time taken to complete tasks relative to minimum realistic time	Rating scale for satisfaction with advanced features
Meets needs for walk-up and use	Percentage of tasks completed successfully at first attempt	Time taken on first attempt to complete task Time spent on help functions	Rate of voluntary use (after system is implemented)
Meets needs for infrequent or intermittent use	Percentage of tasks completed successfully after a specified period of non-use	Time spent re-learning functions Number of persistent errors	Frequency of reuse (after system is implemented)
Learnability	Number of functions learned Percentage of users who manage to learn to a pre-specified criterion	Time spent on help functions Time to learn to criterion	Rating scale for ease of learning

68

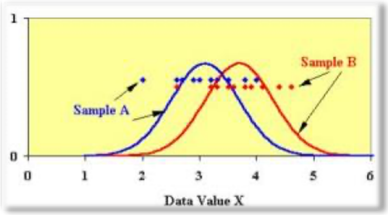
Finding a research design

- Experimental units may be people, but could also be images, samples of sound, videos, etc.
  - How to assign experimental units to levels of the independent variable?
  - Main concerns are realism and generalizability
- Dealing with the ”other factors”---when in doubt, randomize, control, or block

69

Statistics

- Statistical conclusion validity
- Do t-tests or analysis of variance (e.g., in SAS, SPSS, R)
- There are no such things as:
  - Statistics that take away concerns about validity and reliability
  - Significance tests that show the size of an effect or its practical significance
  - Eyeball statistics
- Variability is your enemy



70

Summary of Doing Experiments

- Before doing an experiment:
  - Think about validity and reliability
  - Design so as to rule out alternative hypotheses
  - Consider treatments and measures
- While doing an experiment:
  - Minimize variability
  - Minimize experimenter's influence
- After doing an experiment:
  - Do statistical analysis
  - Discuss ex postfacto validity and sources of error

71

A worked example

- ”It is generally taken for granted that HDR [high-dynamic range] images look better than LDR ones. We put this to test”

Akyüz, A. et al. (2007) ”Do HDR Displays Support LDR Content? A Psychophysical Evaluation”, *ACM Transactions on Graphics*, 26(3), Article 38.

72

73

A worked example II

- Levels of independent variable:
  - HDR capture, HDR display (hdr)
  - HDR capture, LDR display (bila, hist, phot)
  - LDR capture, LDR display (subj, obj)
- 10 scenes

Figure 1: *The test scenes used in our experiment.*

74

A worked example III

- Show 6 images in random order for two seconds each, then presentation of all images in a 2*3 grid
- ”Please indicate which image looks best to you”, no time limit
- Dependent measure is ranking of images
- 22 so-called naïve participants, gender balanced

75

A worked example IV

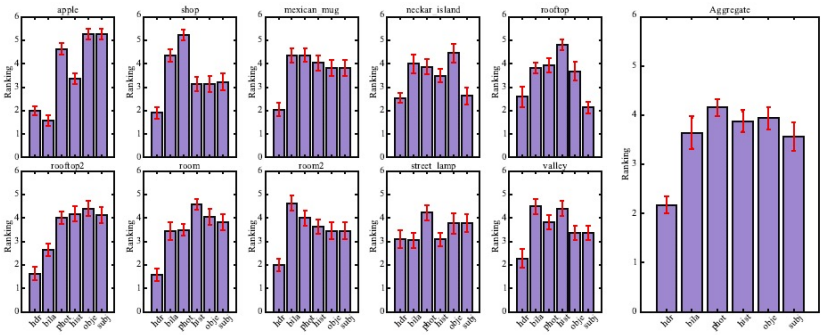


Figure 2: *Mean rankings from Experiment One.* hdr *represents the physically correct HDR image;* bila, phot, hist *represent the bilateral filtering, photographic, and histogram adjustment operators respectively, and* obje *and* subj *represent the objective- and subjective-best images. Left: rankings for the individual scenes. Right: Aggregate result across all scenes. Note that a lower ranking indicates a higher preference. Error bars denote* ± 1 *standard error.*

- Analysis of variance shows $F(5,105) = 20.132$, $p < 0.001$

76

77

Spørgsmål

78

Det fik I med

- Test af brugsvenlighed betaler sig, gør systemer bedre og hjælper til at inddrage brugere
- Et overblik over beslutninger i evaluering
- Tænke-højt studier er et vigtigt redskab

79