Prototype Evaluation II

Human Computer Interaction
CSCI 4620U| SOFE 4850U | CSCI 5540G
Dr. Christopher Collins

Acknowledgement: Parts of this lecture are based on material prepared by Jeremy Bradbury

REVIEW

Goals of Evaluation

- Evaluation in the context of Human-Computer Interaction has 3 main goals:
  - Test the system functionality (and accessibility of the functionality)
  - Test the user's experience interacting with the system
  - Identify specific system problems

Learnability

- Predictability
- Synthesizability
- Familiarity
- Generalizability
- Consistency

Flexibility

- Dialog initiative
- Multi-threading
- Task Migratability
- Substitutivity
- Customizability

Robustness

- Observability
- Recoverability
- Responsiveness
- Task Conformance

Discount Usability Engineering

- Expert Analysis techniques
  - Heuristic Evaluation
  - Cognitive Walkthrough
- Find problems, but not complete, nor ecologically valid
- Formative
  - Help inform design at early stage
- Not summative
  - Cannot make claims about usability of a finished product

Cognitive walkthrough

Required Artifacts

- A system prototype (or specification)
  - Should be as complete as possible
- Description of the task being used for the walkthrough
- The complete, written list of actions necessary to complete the task
- User profile
  - Details about skill level, background, etc.

Cognitive walkthrough

How does an expert conduct a cognitive walkthrough?

- Once the artifacts are gathered by the evaluator they are given to the expert for review
- As the expert walks through each action and move towards completing a task a set of questions must be answered
- Throughout the process the expert will document the procedure and any problems encountered

CW Caution!

- Many evaluators stumble through an interface they don't know, and then evaluate the stumbling process.
- But the aim is to evaluate the optimal sequence and fix all trouble spots that cause a deviation from that sequence.

CW Questions

- Upon completing an action the expert asks himself/herself the following questions:
  - *"Is the effect of the action the same as the users' goal at that point?"*
  - *"Will users see that the action is available?"*
  - *"Once users have found the correct action, will they know it is the one they need?"*
  - *"After the action is taken will users understand the feedback they get?"*

- Benyon, p. 106

CW Experience

Task: Reply to oldest message on Blacboard discussion board “Epic Fails and Epic Wins”

CW Experience

Assumes: default sort (newest first) and default paging (20 per page)

Ideal sequence:

1. Log in to Blackboard
2. Select course
3. Select “discussions”
4. Select “Epic Fails & Wins”
5. Select last page and click green arrow
6. Select message
7. Click “reply”
8. Type something
9. Click “post”

Today

- By the end of today's lecture, you will be able to:
  - Describe the difference between internal and external validity
  - Create a lab experiment and a field study
  - Conduct simple controlled experiments, observations, and think-aloud studies

Benyon text
chapter 10
pp. 232-246

Evaluating HCI

PARTICIPANT-BASED EVALUATION

Evaluation with Real Participants

- We can engage real primary stakeholders as participants in a variety of evaluation techniques
- These techniques, including laboratory studies and field studies, use qualitative and quantitative methods
- Can reveal usability and usefulness problems with early prototypes (paper) through to late-stage software (alpha/beta release)

User Participation

- Both laboratory and field studies can be costly
  - Field studies involve interruption to work day and onsite evaluation
  - Laboratory studies also interrupt participants work day and require them to travel to the lab
- There are benefits and drawbacks to both which we will now discuss

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

Laboratory Studies

Advantages

- Involves the user participating in a controlled environment – no distractions
- Can be equipped with monitoring equipment to record the study
  - Video, audio, keylogging, physiological

http://magazin.unic.com/wp-content/uploads/2011/06/SBB_Usability-Testing.jpg

Laboratory Studies

Disadvantages

- "Unnatural environment"
- Not as useful for group scenarios that involve collaboration
- Often involves non-typical people as participants (e.g. students stand in for real stakeholders)

Field Studies

Advantages

- "Natural" environment
- Can identify the effect of other work activities on use of system
- Can study long-term activities (spanning more than a few hours)

Field Studies

Disadvantages

- Environment is not controlled – constant distractions can affect the outcome
- **Example:** text messages, meetings, etc.

Lab ≠ Quant; Field ≠ Qual

- Lab:
  - Quantitative example: have participant click a button 20 times and measure speed and accuracy
  - Qualitative example: ask participant to complete a purchase on the new website and observe
- Field:
  - Quantitative example: observe through software logging the way visitors interact with wayfinding kiosk over time
  - Qualitative example: observe several individuals using the kiosk and approach them to participate in an interview

Quantitative Methods

- In an experiment it is important to:
  - select the appropriate participants
  - Define the independent and dependent variables
    - Independent = manipulated
    - Dependent = measured
  - Have a clear hypothesis
  - Have a well thought out experimental design
  - Use appropriate statistical measures

Quantitative methods are also called empirical methods.

Confounds

- Confounding variables get in the way:
  - learning effects / order effects
  - the effects of different tasks,
  - the effects of different background knowledge, etc.
- You want to ensure a balanced and clear relationship between independent and dependent variables so that you can be sure you are looking at the relationship between them and nothing else.

Quantitative Methods

- Controlled experiments are appropriate where the designer is interested in particular features of a design, perhaps comparing one design to another to see which is better.
- In order to do this with any certainty the experiment needs to be carefully designed and run.

Common Usability Metrics

Table 10.3 Common usability metrics

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

Source: ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs), extract of Table B.2

Lab Studies: What are you testing?

Low level

- Comparing two interaction designs to see which is faster/more accurate
  - Example: mouse vs. pen to tap a button
- Simple, repeated tasks
- Several conditions compared

Controlled study with repeated tasks completion; usually automated measures and questionnaire at end

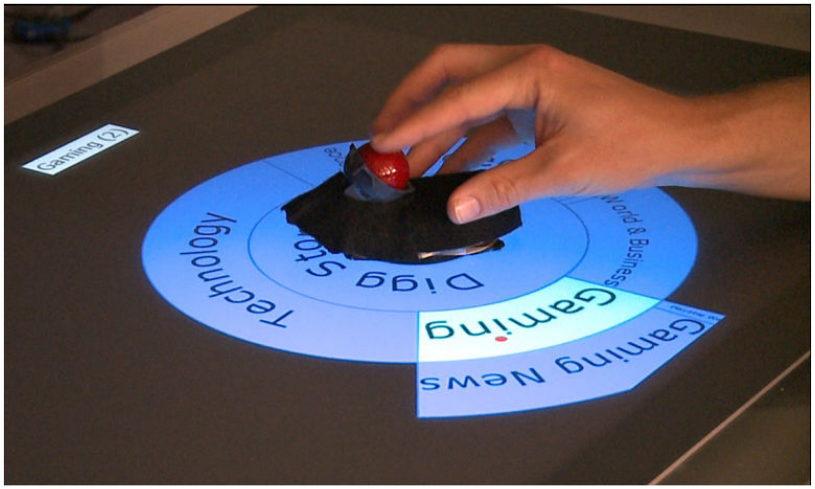
High level

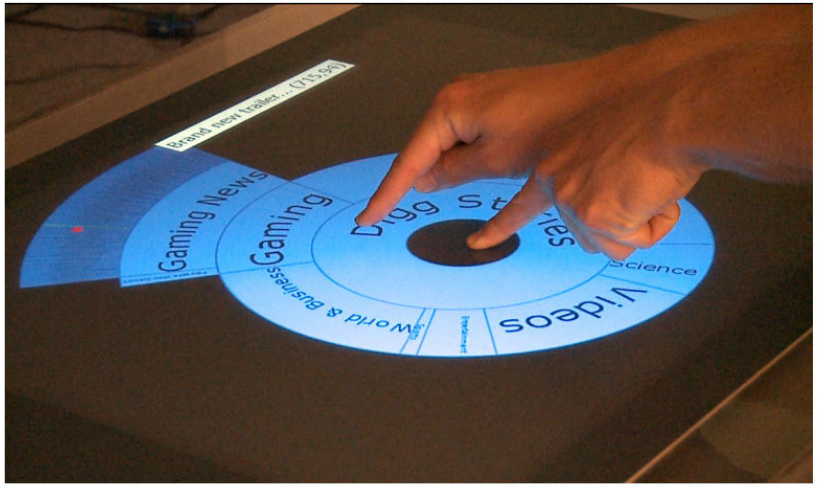
- Evaluate a new interface design or task sequence
  - Example: is our interactive cooking tutorial website effective and easy to use?
- Longer tasks
- Often one condition

Full usability study with tasks and methods such as think aloud; qualitative focus but also uses quantitative measures

Let’s design a controlled experiment!

- Compare two indirect pointing methods – two finger and track ball; which is better for selection?
- This is a low level experiment, not a usability study.





USABILITY STUDY BASICS

Usability Testing Overview

- Running Usability Tests
  - Roles & Responsibilities
  - Methodology
  - Preparing & helping participants
  - Capturing data
- Usually yields surprising results
- Produces answers to questions not asked

Usability Test Participants' Roles

- Test subject ("user"/ "participant") (or two)
- Facilitator
- Computer (optional - manage digital or paper prototype)
- Observer(s)

Facilitator Responsibilities

- Before participant (*p*) arrives, brief in-room observers re rules
- Greet *p* outside, brief them, obtain consent, pay honorarium
- Escort *p* inside & introduce to computer & observers
- Explain testing protocol
- Facilitate each task, interacting when necessary
- Manage time on each task, cover all key points
- (optional) Facilitate short Q&A session with *p*
- End session on time
- Thank *p* & escort them out
- Debrief observers, list issues
- (possibly) Make changes before next test.

Consent Form

- Description of what participant will do
- What is being tested (interface, not participant)
- Any risks to physical or psychological well-being (presence of observers)
- Length of session, & right to withdraw
- Data collected and its intended usage
- Possibly: Non Disclosure Agreement NDA

Metaphors for Facilitator Roles

- Flight Attendant
  - safeguard participants' well-being (physical, psychological, emotional)
  - reduce users' stress, give help & praise as appropriate
- Sportscaster
  - help observers get maximum useful info from test
  - provide play-by-play & remind users to think aloud
  - encourage users to ask questions but don't answer them
  - use users' vocabulary & open ended questions
  - watch for nonspecific utterances & body language
  - learn to Shut Up, let users decide when they're done
  - end task early if appropriate
- Scientist - preserve integrity of data
  - Maintain objectivity, qualitative observations
  - Take notes (hard for beginners) for later discussion
  - Beware of biasing results: hint, smile, body language

Tips for New Facilitators

- Use a a checklist!
- Try not to read from scripts!
- Practice out loud
- Seek feedback to improve your skills
- Strive for progress, not perfection - mistakes happen during testing, get over it & move on

Co-Discovery (Two-User) Testing

Benefits

- natural conversation & problem solving during test
- helps users relaxed while being observed
- easier for facilitator, less need to talk & provide positive feedback
- more data generated by two users (not independent 1.5 of single test)
- easier scheduling - less no shows & cancellations

Drawbacks

- discrepancy in experience or confidence, harder to recruit balanced teams
- peer pressure concern about looking bad
- slower pace, more discussion, slower task times
- dominant personality, difficulty sharing

Friends or Strangers

- strangers do not share “communication shorthand”, easier to observe
- product dependent choice - family product, for example
- users more comfortable working with friend

Pretest Briefing (Outside Test Room)

Topic	Checklist
Greeting & introduction	Welcome user, introduce self & company, describe product being tested & paper prototyping, get consent form signed, explain session goals
User role	Introduce two users (if co-discovery), explain whats expected of them, remind them of their qualifications(you are here…)
Social concerns	Mention in room observers, explain testing interface not you, reassure what happens if they have difficulty, express appreciation
Set expectations	Describe unfinished prototype (as a benefit), explain evolving design, explain data recording
Paperwork & administrivia	Get consent form signed, pay users, escort into test room

Based on www.paperprototyping.com/downloads/Table9.1-2_pretest.pdf

Introducing the Test (Inside Test Room)

Topic	Checklist
Introductions	Show users seat, ask observers to introduce themselves, introduce users, users answer 2-3 background questions
Prototype Orientation	Explain what they're looking at (first screen), (optional) to novice users explain common parts of interface like browser buttons
Interactions w Prototype & Computer	Explain how to interact, discuss think-aloud & questions, (optional) explain limited interface functionality, show available help or print manuals
User in Charge	Remind about testing interface not user, confirm ending time, remind of break & end possibility
Begin first task	Give first task, clarify if confusing,(if necessary) prompt to begin interacting

Based on www.paperprototyping.com/downloads/Table9.1-2_pretest.pdf

Testing Paper Prototypes: "Computer" Participant Behavior

- Accurately reflect users' inputs
- Wait for users, avoid anticipating, don't get next screen too early
- Avoid conversation
  - if user needs help reading the screen, assist with what user would see, but NOT why
  - Avoid explaining & using BECAUSE
- Provide oral error messages
  but only as the system would, no extra context

Rules for Usability Test Observers

- Stay for the entire test (& turn off phone)
- Remain silent while participants are working
- Be conscious of your body language
- Don’t reveal how many tasks you have
- Avoid “design questions” - “where would you like to see…” Try to understand the problem instead
- Respect participants & confidentiality of their data
- No helping unless requested by facilitator

Getting Participants Unstuck

- Not if but *how soon & how badly* will participants get stuck during a test?
- Most valuable observation: participant having difficulty
- Resist temptation to answer questions!
- Once you explain something, you forever lose an opportunity to understand the problem:
  - participants get embarrassed of their ignorance
  - participants can't reconstruct their thoughts once they have the solution

“Once you start explaining the interface, you no longer have a usability test but a very expensive training session!”

Prof. R. Wolfe, DePaul University
in Snyder, 2003, pg. 227

Example: Getting Participants Unstuck

Facilitator	Participant
• What are you trying right now?	• I want to transfer this call to Mike in Accounting.
• What do you think is the next step?	• I want to dial 5385, but not lose this guy I'm talking to.
• (Small hint) Do you see anything that might help you?	• I'm not sure, if I dial the extension, won't it just beep?
• (Big hint) What do you think the Flash button does?	• I was wondering about that, but I was afraid it would hang up.

Here, facilitator verifies that:

- participant is trying to do what you thought
- shares understanding of how to do it
- can find and use specific function or control needed

PROBLEM: Insufficient information about how Flash button works

Capture & Analysis of Data

- Capture
  - keystroke logs & screen capture
  - task completion logs – accuracy/time
  - video and audio recording
  - physiological – galvanic skin response, heart rate
  - note taking
- Analysis
  - quantitative
  - qualitative

Capturing Data: Automated Measures

- Keystroke logging
  - Precise participant behaviour
  - Times to carry out actions
  - Participant errors
- Audio and video recordings
  - Can't observe and record all behaviour in real-time
  - Preserve behaviour for review
  - Non-verbal behaviour
  - Behaviour in context

Capture and Analysis of Data

- Data extraction and analysis
  - Review of video and audio recordings, annotation and analysis
  - Quantitative data, e.g., classifying events, counting events
- Interaction analysis
  - Interaction analysis..."to uncover the regularity and efficacy of peoples' relations with each other and their use of the resources that their environment affords"
  - Used in ethnographic observations of real work contexts, and in analysis of user testing or other lab studies
  - Qualitative data, e.g., quotes, non-verbal behaviour descriptions

Capturing Data: Note Taking

- Helps focus on users and test
- Note confusion with terminology, concepts
- Surprising behavior & observations important
- Things that work well need recording
- Record computer generated error messages
- Take note of users' navigation path & focus
- User quotes are very valuable!

Capturing Data: Notation Types

- Observations - objective, factual statement
  - “user paused for a couple of seconds”
- Inferences - conclusions based on observations & assumptions
  - “paused because: user didn’t know format of data entry/process steps/tried to remember something”
- Opinions - use “should” & “need”
  - “need clearer data entry format instructions”

Capturing Data: Notation Types

- Try to make only observations during the experiment
- A good facilitator can turn an inference into an observation with a question, e.g. “What is going through your mind right now?”
- Inferences and Opinions can be formed with the complete observational evidence starting during debriefing meeting and in further analysis.

> “Inferences and opinions in your notes are ‘argument seeds,’ likely to sprout into disagreement later.”
>
> Snyder, 2003, pg. 244

Bias in Usability Tests

- Users who don’t fit the profile: tech expertise varies
- Tasks or instructions not matched to chaos of reality
- Test setting: Usability lab’s not a natural work env’t
- Test machine: Less conflicts & disruptions, speed
- Facilitator: Leading questions, hints, social setting
- Methodology: Think aloud, no interruptions, video
- Data Analysis & Reporting: Inter-rater reliability
- Observers: Users behave differently when observed
- Paper Prototype: Slow speed discourages exploration

http://jennycham.co.uk/wp-content/uploads/2011/08/20091122125022548l.jpg

METHODOLOGICAL VARIATIONS

Qualitative Methodologies

- Think aloud
  - Participant talks through a task aloud under observation
- Retrospective think aloud
  - Participant watches video of task completion and explains actions and thoughts
- Cooperative evaluation
  - Participant talks through a task aloud
  - Evaluator can ask questions
  - User can ask for more instructions or help
- Conceptual model evaluation
  - Participant views static images and explains components

http://kilon.org/blog/2012/10/usability-testing-in-code-review/

Think Aloud Method

- Participants 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
  - wishes for improvements
- Gives insight into what the user is thinking
- Most widely used evaluation method in industry
- Problems:
  -
  -
  -

Think Aloud Studies

- Participants 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
  - wishes for improvements
- Gives insight into what the user is thinking
- Most widely used evaluation method in industry
- Problems:
  - may alter the way users do the task
  - unnatural (awkward and uncomfortable)
  - hard to talk if they are concentrating

Think Aloud Method

- Create a task
- Explain the system concept and task goal
- Ask the participant to use the system to achieve the goal while *thinking aloud*
- Evaluator takes notes, provides minimal help
- Step in only when it is clear the participant is stuck, then make a note of where it happened

Retrospective Think Aloud

- The participant can comment as the observation data is replayed
- Pros: get the interpretation directly from participant, does not disrupt task completion
- Cons: participant may not remember or may accidentally misinterpret their own actions

Retrospective Think Aloud is also called a Post Task Walkthrough

Retrospective Think Aloud Method

- Perform an usability test
- Create a video record of it
- Have participant view the video or other captured data and comment on what they did
  - clarify events that occurred during system use
  - excellent for grounding a post-test interview
  - avoids erroneous reconstruction
  - participants often offer concrete suggestions

Cooperative Evaluation

- Talk to participant as he or she works
- Alternative to Think Aloud
- Very helpful in paper prototyping tests
- Ask questions, encourage to elaborate, remain neutral

Usability Testing with a Paper Prototype

www.blueducklabs.com

Conceptual Model Evaluation

- 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
- 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

Data Interpretation: Coding

- Another possibility is protocol coding
  - The experimenter watches video / listens to audio and carefully counts and timestamps events of interest

Coding sheet example...

- tracking a person's use of an editor

	General actions			Graph editing			Errors	
Time	text editing	scrolling	image editing	new node	delete node	modify node	correct error	miss error
09:00	X							
09:02				X				
09:05							X	
09:10					X			
09:13								

most common process in our study.

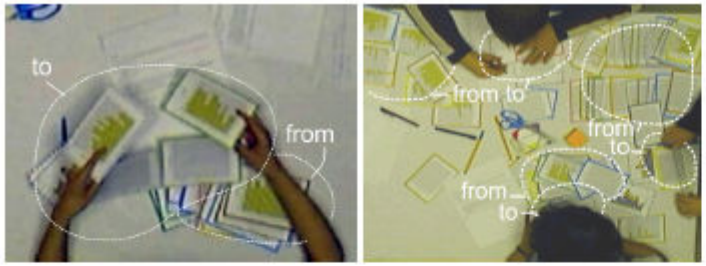


Figure 5. Chart organization during selection depending on their intended usage. *Left*: a participant selected four cards for comparison placing them side by side in her hand. *Right*: three participants selected individual charts and placed them in the center of their workspace to measure a specific value.

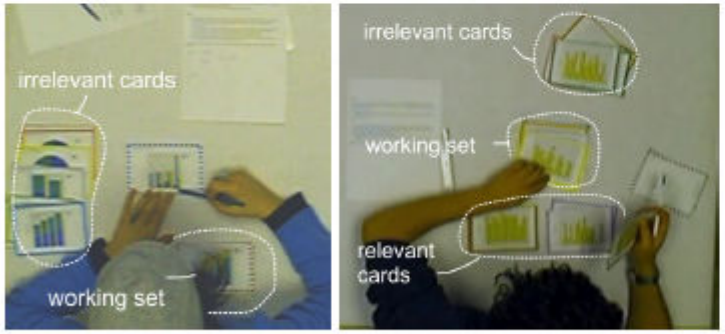


Figure 6. Changing categorization during selection. *Left*: a participant placed irrelevant cards to her left and picks single cards to operate on from the working set. *Right*: a participant picked out relevant cards, placed them close to himself, and put irrelevant cards in a pile further away.

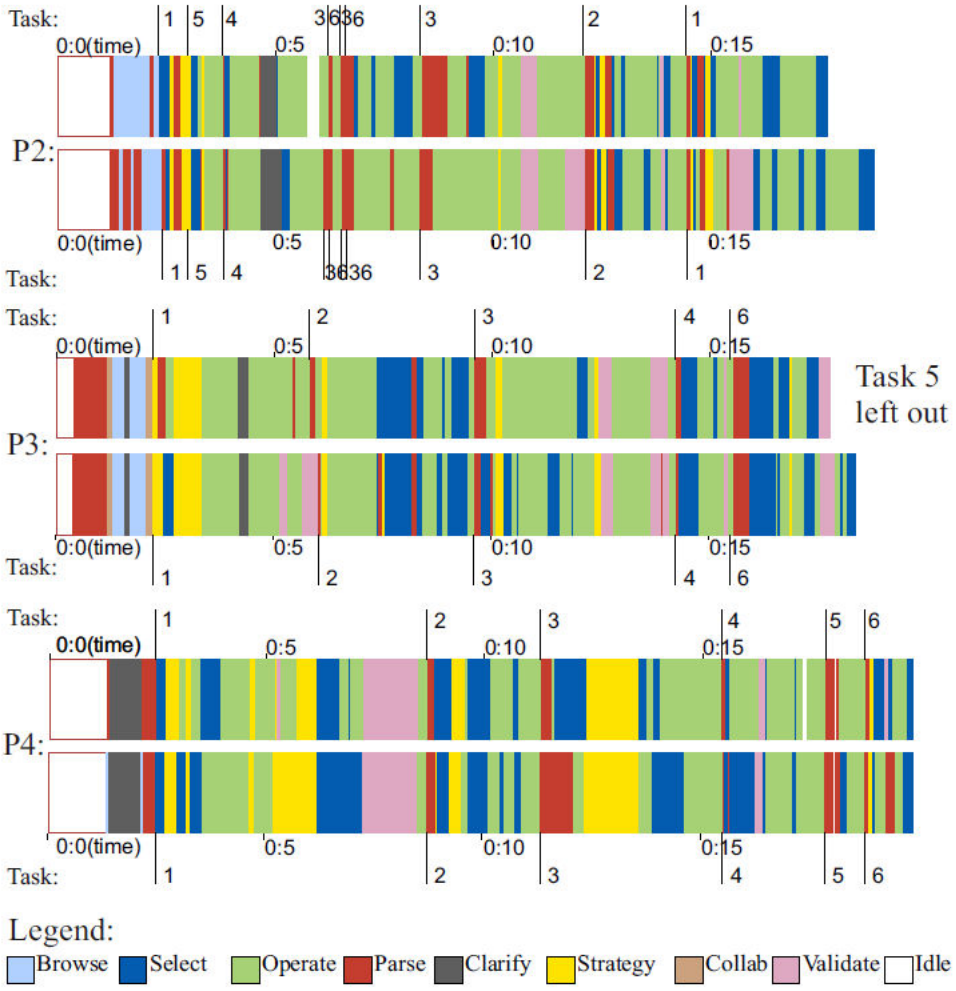


Figure 8. Temporal sequence of processes for three pairs during one complete scenario.

Petra Isenberg, Anthony Tang, and Sheelagh Carpendale. An Exploratory Study of Visual Information Analysis. CHI 2008.

Query Techniques

- There are two main query techniques we have already considered
  - Interviews
    - Typically structured but with some flexibility
  - Questionnaires
    - Needs to be very well designed
    - Contains different kinds of questions:
      - General questions
      - Open-ended questions
      - Scalar questions

Critical incidence interviews

- Ask about the biggest problems
- 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

http://nform.com/cards/usability-testing

Query Techniques

- Where previously you investigated existing contexts, routines, interfaces, you can also use query techniques to elicit responses about a prototype design

Continuous Evaluation

- Monitor systems in actual use
  - Usage logging built in to the system
  - Usually late stages of development
    - beta releases, delivered system
  - Fix problems in next release
- Feedback via in-program forms
  - Participants can provide feedback to designers while using the system (not only for crashes!)
  - Best combined with trouble-shooting facility
    - Participants always get a response (solution?) to their gripes

On-Site Continuous Evaluation

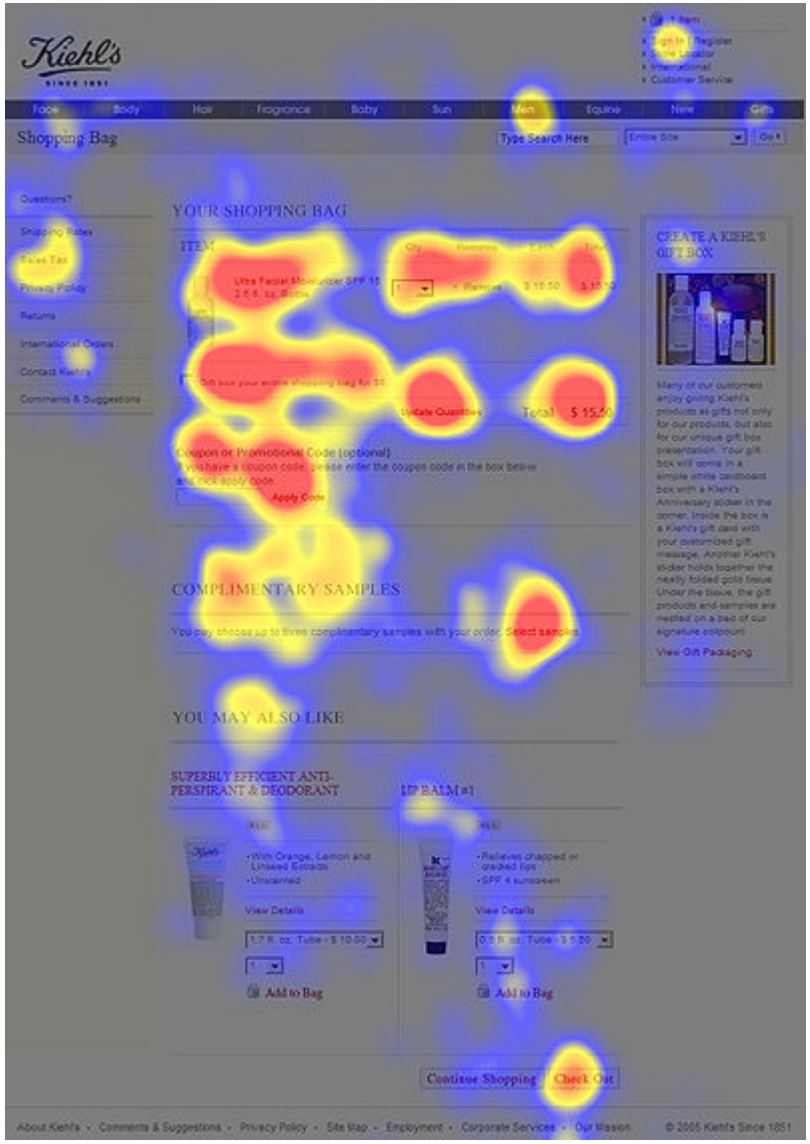
- Case/field studies
  - careful study of “system usage” at the site
  - good for seeing “real life” use
  - external observer monitors behavior
  - site visits on a regular interval

Eye-tracking

- Eye-movement tracking (or ‘eye tracking’) can show participants’ changing focus on different areas of the screen
- Can indicate which features of a user interface have attracted attention, and in which order, or capture larger-scale gaze patterns
- Popular with web site designers as it can be used to highlight which parts of the page are most looked at, so-called ‘hot spots’ and which are missed altogether
- Head-mounted or attached to computer monitors
- Software provides maps of *fixations* on the screen

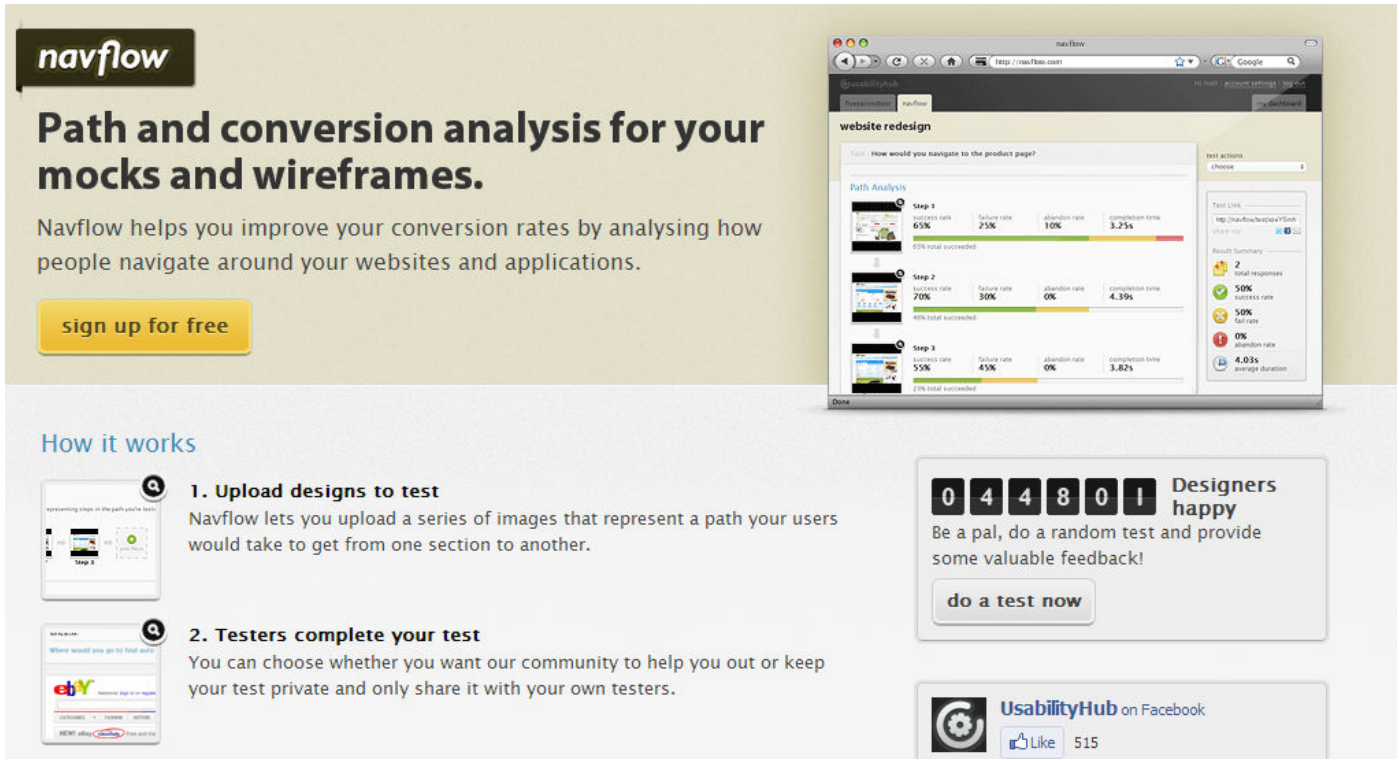

tobii


tobii

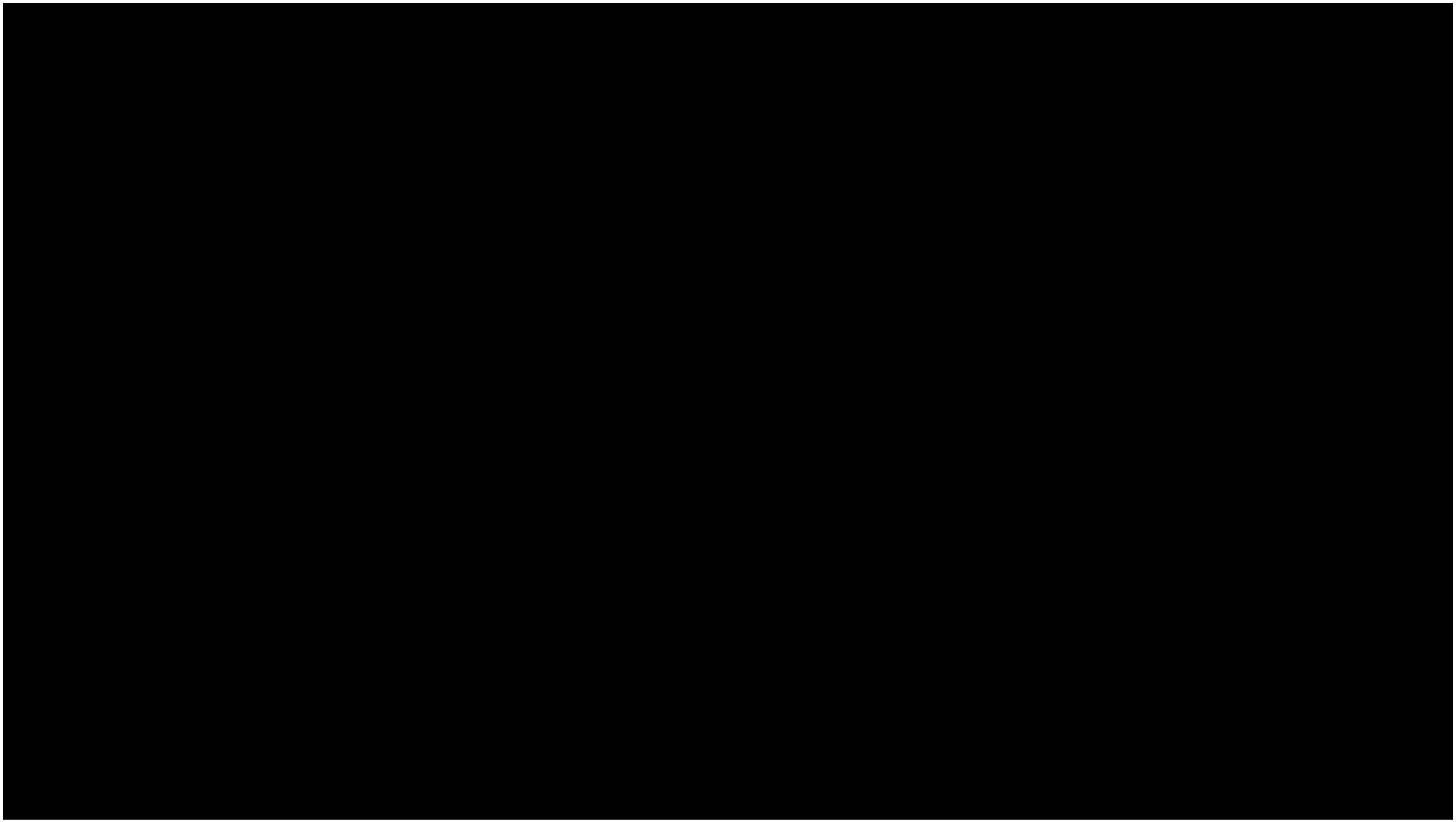

Kiehl's
Face
Body
Hair
Fragrance
Baby
Sun
Men
Equine
New
Gifts
Shopping Bag
Type Search Here
Go
YOUR SHOPPING BAG
ITEM
Total $ 15.50
Coupon or Promotional Code (optional)
Apply Code
COMPLIMENTARY SAMPLES
YOU MAY ALSO LIKE
SUPERBLY EFFICIENT ANTI-PERSPIRANT & DEODORANT
LIP BALM #1
Add to Bag
Add to Bag
Continue Shopping
Check Out
CREATE A KIEHL'S GIFT BOX
View Gift Packaging
About Kiehl's
Comments & Suggestions
Privacy Policy
Site Map
Employment
Corporate Services
Our Mission
© 2005 Kiehl's Since 1851

New: Crowdsourcing Usability

- http://fivesecondtest.com/
- http://navflow.com/



Fruit Usability



Summary

- Today we introduced:
  - Qualitative and Quantitative methods involving participants

Your Action Items

- Watch Lecture 17 video if you have not – needed for the lab this week!
- Finalize part 3b. Contact us for help sooner rather than later!

Ongoing Course Evaluation

- Please complete lecture feedback form online.