



Intro to Evaluation & Analytical Evaluation



Agenda

- Announcements
- Lecture:
 - Evaluation Overview
 - Analytical Evaluation
- Evaluation Activity: Heuristic Evaluation
- Next class



Announcements, Questions

- P2 due now
- Work on P3 Prototypes (due 12/4)
 - As you are working on your prototypes, be thinking about your evaluation plan
 - The turn time between P3 and P4 is SHORT, because of our delay in turning in P1
- Questions?



P3: Prototyping

- Create an interactive, hi-fi prototype of your design idea(s) that shows some details of the interaction
 - Method of prototyping is up to you
 - You may use multiple methods if you need to do so to convey aspects of your prototype, but at least one of these must be hi-fi
 - You do not need to create more than one
- Deliverables
 - Demo of your prototype during class on 12/4 (5mins)
 - Short report on the Q&A from class (due 12/6)



P4 Description

- Prepare a "trade show" presentation
 - To be given to anyone who wants to come on 12/10 at noon in DBH 5011
 - Food!!!!
- Write your eval plan & project summary – Due 12/14 @5pm to EEE
 - Now that you've come up with your design solution, how would you evaluate it? Try this out with a few people, fix your eval plan as necessary
 - Describe how you might change your design and your eval plan based on your experiences at the trade show and in your pilot evals

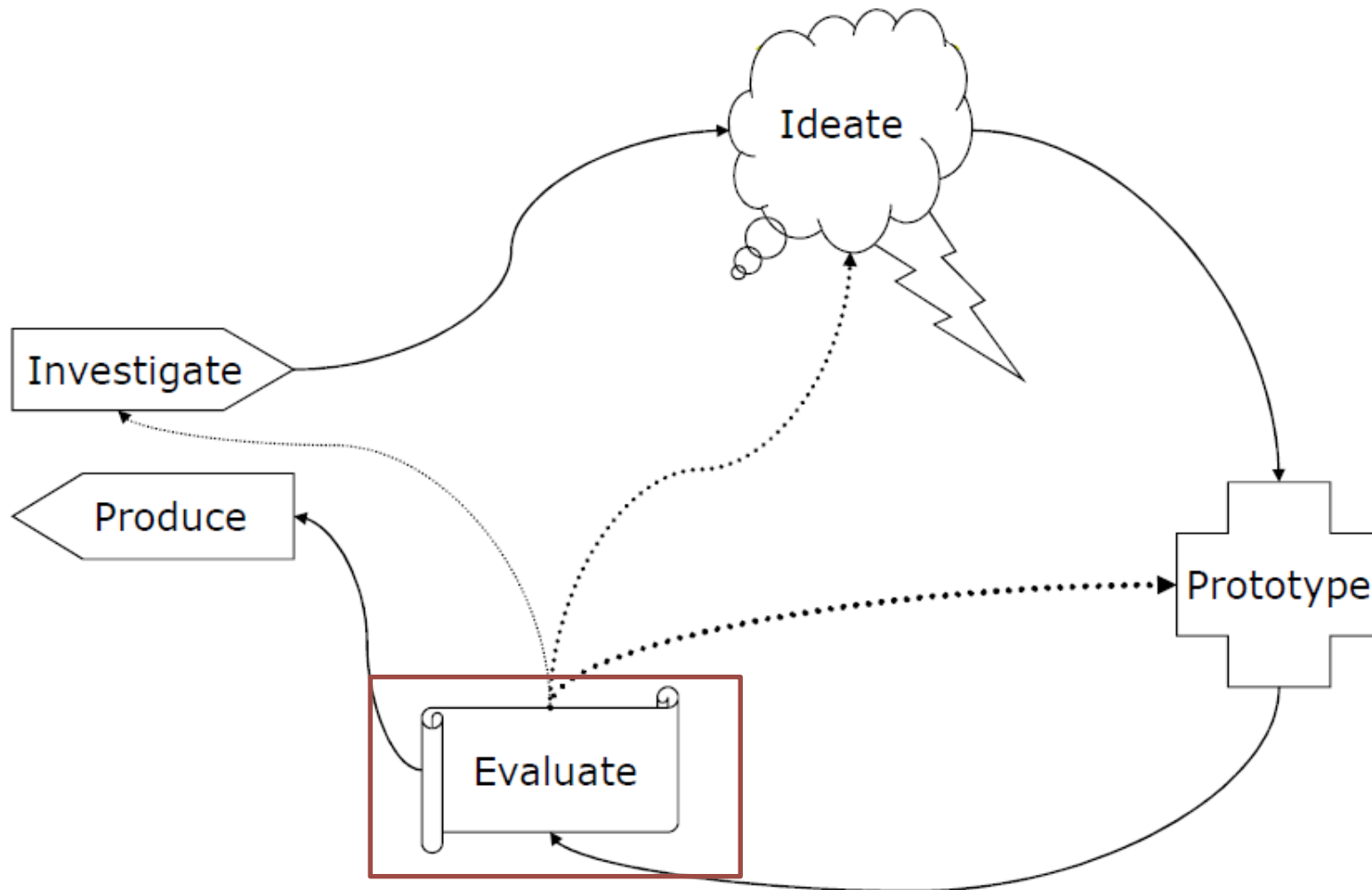




LECTURE – EVALUATION OVERVIEW



In the UCD Process



Many Approaches to Evaluation

Usability goals:

- Effectiveness
- Efficiency
- Safety
- Utility
- Learnability
- Memorability

User experience goals

- Satisfying
- Pleasurable
- Rewarding
- Fun
- Provocative
- ...

***Consider your project. What are your usability goals?
What are your user experience goals? How would you
define and operationalize these goals?***



Why and Where to Evaluate

- Why?
 - Feedback on design directions and ideas
 - Discover major issues
 - (Help to) resolve disagreements
- Where?
 - In laboratory (controlled)
 - In natural settings (uncontrolled)



When to Evaluate

- Early design of an artifact
- Evaluation of working prototype
- Refining or maintaining a product
- Competitive comparison between two products
- Exploring an new design concept
- Demonstrate performance for a procurement contract



Types of Evaluation

- Analytic (design judgment – users not involved)
 - Often called “discount evaluations”
 - Standards enforcement
 - Heuristic evaluations
 - Cognitive walkthroughs
- Empirical (involves users)
 - Usability testing
 - Field studies
 - Click-through studies





Part One: ANALYTICAL EVALUATION



Analytical Evaluation

- Cognitive Walkthrough
 - Have experts analyze your prototype in a detailed way to understand how users will understand it
 - Best for understanding novel use, not expert use
 - http://en.wikipedia.org/wiki/Cognitive_walkthrough
- Heuristic Evaluation



Cognitive Walkthrough

- Uses a small number of HCI experts to evaluate a design for ease of learning, especially via exploration
- Analogy to code walkthrough (Polson, Lewis, *et al.* at UC Boulder)



Cognitive Walkthrough

- Requires prototype or fairly detailed description
- Requires a description of the user task to be analyzed
- Requires a complete, written list of actions necessary to complete the task
- Requires an indication of who the users are and their assumed knowledge



Procedure

- Define required inputs
- Walk through action sequences for task
- Record critical information & obtain believability story



Inputs

(1) Select Interaction Task

- Task should be a one that would be common or typical for a potential user
- Should be representative of what users would want to do with the system

(2) Define interaction action sequence

- Tasks should be broken down until any further division yields obvious subtasks
- E.g., type “run” at prompt NOT type “r”; type “u” etc.



Inputs

(3) Identify the users

- Educate the HCI experts on the domain knowledge, experience, and characteristics of the user
- Give evaluators a perspective from which to evaluate the system

(4) Prototype

- Need not be functional but...
- Must be at a level of detail where any action necessary to complete the task is defined



Doing the walkthrough

- Address each step of task sequence in turn
- Formulate a believability story
 - Answer 4 questions
 - Include justification for each answer based on the interface, knowledge of HCI, and understanding of users



Question 1: *Will the user be trying to produce whatever effect the action has?*

- Common supporting evidence
 - It is part of their original task
 - They have experience using the system
 - The system tells them to do it
- No supporting evidence?
 - Construct a failure story
 - Why would the user *not* be trying to do this?



Question 2: *Will the user be able to notice that the correct action is available?*

- Common supporting evidence
 - Known through experience
 - Visible device, such as a button
 - Visible representation of an action, such as a menu entry
- No supporting evidence?
 - Why would the user *not* notice such an action is available?



Question 3: *Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?*

- Common supporting evidence
 - Based on past experience with similar interactions
 - The interface provides a prompt or label that connects the action to what he/she is trying to do
 - All other actions look wrong
- If not, why not?



Question 4: *After the action is taken, will the user understand the feedback given?*

- Common supporting evidence
 - Past experience with similar interactions
 - Recognizing a connection between the system response and what the user was trying to do
- If not, why not?



Analytical Evaluation

- Cognitive Walkthrough
- Heuristic Evaluation
 - Have usability experts go through your prototype to uncover common usability problems



Heuristic Evaluation

- Developed by Jakob Nielsen
- Helps find usability problems in a UI design
- Small set (3-5) of evaluators examine UI
 - independently check for compliance with usability principles (“heuristics”)
 - different evaluators will find different problems
 - evaluators only communicate afterwards
 - findings are then aggregated
- Can perform on working UI or on sketches



Heuristic Evaluation Process

- Evaluators go through UI several times
 - inspect various dialogue elements
 - compare with list of usability principles
 - consider other principles/results that come to mind
- Usability principles
 - Nielsen's "heuristics"
 - supplementary list of category-specific heuristics
 - competitive analysis & user testing of existing products
- Use violations to redesign/fix problems



Heuristics (Nielsen, 1994)

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation



Phases of Heuristic Evaluation

1) Pre-evaluation training

- give evaluators needed domain knowledge & information on the scenario

2) Evaluation

- individuals evaluates UI & makes list of problems

3) Severity rating

- determine how severe each problem is

4) Aggregation

- group meets & aggregates problems (w/ ratings)

5) Debriefing

- discuss the outcome with design team



How to Perform Evaluation

- At least two passes for each evaluator (3-5 people)
 - first to get feel for flow and scope of system
 - second to focus on specific elements
- If system is walk-up-and-use or evaluators are domain experts, no assistance needed
 - otherwise might supply evaluators with scenarios
- Each evaluator produces list of problems
 - explain why with reference to heuristic or other information
 - be specific & list each problem separately



Example Errors from Evaluators

- Can't copy info from one window to another
 - violates “Minimize the users’ memory load” (H3)
 - fix: allow copying
- Typography uses different fonts in 3 dialog boxes
 - violates “Consistency and standards” (H4)
 - slows users down
 - probably wouldn’ t be found by user testing
 - fix: pick a single format for entire interface



Severity Rating

- Used to allocate resources to fix problems
- Estimates of need for more usability efforts
- Combination of
 - frequency
 - impact
 - persistence (one time or repeating)
- Should be calculated after all evals. are in
- Should be done independently by all judges



Severity Ratings (cont.)

- 0 - don't agree that this is a usability problem
- 1 - cosmetic problem
- 2 - minor usability problem
- 3 - major usability problem; important to fix
- 4 - usability catastrophe; imperative to fix



Debriefing

- Conduct with evaluators, observers, and development team members
- Discuss general characteristics of UI
- Suggest potential improvements to address major usability problems
- Dev. team rates how hard things are to fix
- Make it a brainstorming session
 - little criticism until end of session



Severity Ratings Example

1. [H4 Consistency] [Severity 3]

The interface used the string "Save" on the first screen for saving the user's file, but used the string "Write file" on the second screen. Users may be confused by this different terminology for the same function.



HE vs. User Testing

- HE is much faster
 - 1-2 hours each evaluator vs. days-weeks
- HE doesn't require interpreting user's actions
- User testing is far more accurate (by def.)
 - takes into account actual users and tasks
 - HE may miss problems & find “false positives”
- Good to alternate between HE & user testing
 - find different problems
 - Don't waste participants



Class Activity: Heuristic Evaluation

- Electronic voting machine
 - Download prototype (swf file on website)
 - Download form (xls file on website)
 - Nielsen's heuristics:
 - http://www.useit.com/papers/heuristic/heuristic_list.html
- Use form and Nielsen's 1994 heuristics to evaluate the voting interface in groups of 3 or 4



Next Class

- Empirical Evaluation
- Upcoming Work
 - A3 due on Tuesday when you get back (last individual assignment!)
 - R8 due Thursday (last reading reflection!)
 - P3 demos 12/4

