

# **Evaluation**

Assoc Prof. Jeannie S. Lee

# Overview

- What is evaluation?
- Goals of evaluation
- Formative vs. Summative evaluation
- Evaluation methods
  - Empirical
  - Predictive
    - Heuristic Evaluation
    - Schneiderman's 8 Golden Rules
    - Cognitive Walkthrough
- DECIDE framework

# Evaluation

Gather data about the usability of a design for a particular activity by a specified group of users within a specified environment

# Goals of Evaluation

- Assess extent of system's functionality
- Assess effect of interface on user
- Identify specific problems with system

# Formative vs. Summative

- Formative evaluation
  - As project is forming
  - All through the lifecycle
  - Early, continuous. Iterative
  - “Evaluating the design”
- Summative evaluation
  - After a system has been finished
  - Make judgments about final item
  - “Evaluating the implementation”

# Multiple ways to evaluate

**Empirical** Assess with real users

**Formal** Models and formulas to calculate measures

**Automated** Software measures

**Predictive/Critique** Expertise and heuristic feedback

# Some Evaluation Methods

- Empirical evaluation
  - Lab studies
  - Field studies
  - “Think Aloud”
- Predictive evaluation
  - Heuristic Evaluation
  - Cognitive Walkthrough
  - Pluralistic Walkthrough (not covered)

# Empirical Evaluation

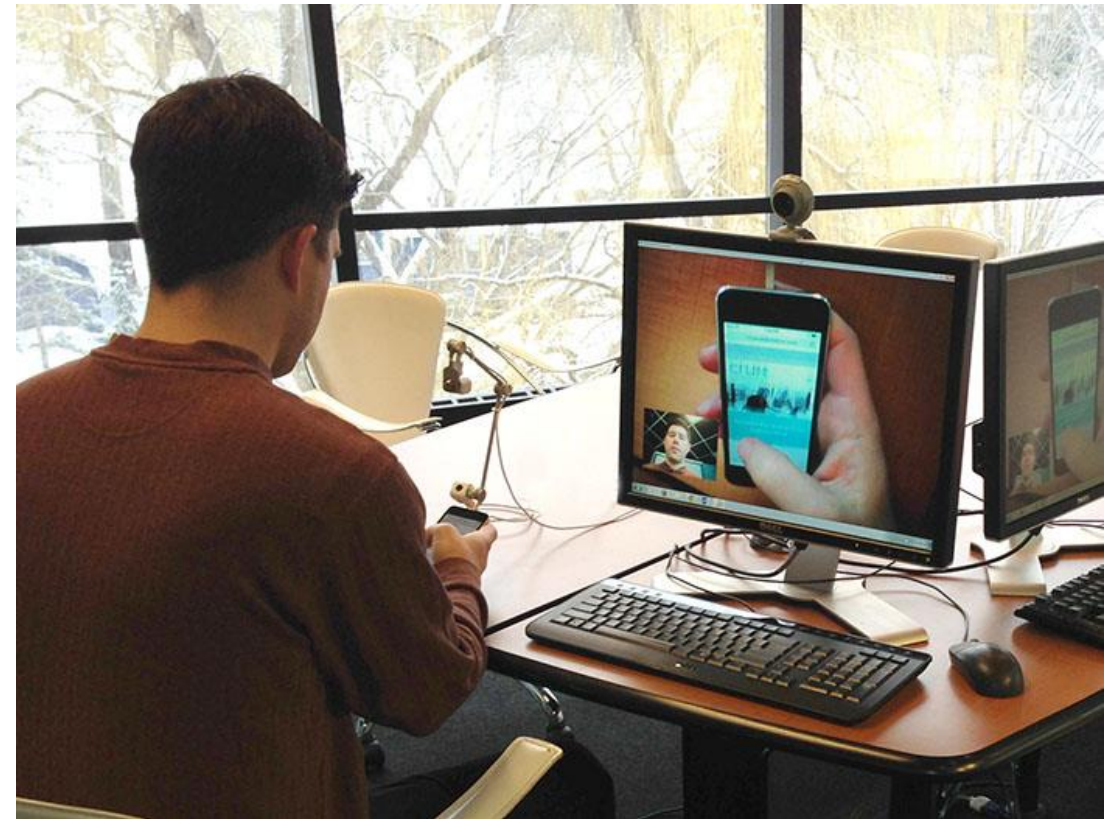
- Verifiable or provable by means of observation or experiment (based on empirical evidence)
- Qualitative or Quantitative
- Conducted with real users
- Earlier in the design process: testing paper prototypes in the lab
- Later in the design process: Field studies, out in the wild, usability testing



# Empirical Evaluation

## Lab studies (quantitative results)

- Typically in a closed, lab setting
- Manipulate independent variables to see effect on dependent variables
- Compare experimental group with the control group
- E.g. new UI vs. existing UI
- Pros: Replicable
- Cons: Expensive, requires real users and lab



# Empirical Evaluation

## Field studies (qualitative results)

- Observation occurs in “real life” setting
- Natural environment
- Watch process over time
- Pros:
  - “Ecologically valid”
  - Cheap, quick, less training required
- Cons:
  - Not reproducible; user-specific results
  - Not quantitative (how much better is it...?)



# User Tasks

- Give user some tasks to complete
  - Tasks related to objectives/goals
  - Objective may have one or more tasks
- Tasks have a description
  - Task user needs to complete
  - Input and expected output
  - Normally included on an information sheet
- Decide on measures and questions
  - Measures to be made during the task: what to measure, how to measure?
    - Time taken to complete the task
    - Number of errors made
    - Number of users completing the task successfully
    - Number of time users had to see information sheet to complete the task

# Think Aloud Protocol

- Ask users to think aloud while completing the tasks
  - Speak loud the steps to complete the tasks
  - Speak up any issues they are facing
  - Speak about confusion they are encountering
- Evaluator role:
  - Conduct observation of the user
  - Take notes
    - E.g. of problems faced, confusion and types of errors made
  - Record video or audio
  - Analyze data later

Useful to know what users are thinking, while the task is being completed

# Think Aloud Protocol

- May need to facilitate user responsiveness
- Possible prompts:
  - “Tell me what you are thinking.”
  - “Tell me what you are trying to do.”
  - “Are you looking for something? What?”
  - “What did you expect to happen just now?”
  - “What do you mean by that?”

Do not help users to do the tasks or give solutions. Want to understand the problems and issues users are facing.



A man with dark hair and a beard, wearing a blue button-down shirt and a black wristwatch, is operating a large industrial machine. He is looking at a touchscreen display on the machine, which shows a graphical interface with various colored buttons and text. The machine is white and black. In the background, there is a red Toyota forklift, metal shelving units with various items, and a large industrial building with a high ceiling and bright lights.

...and, uh, I need to make sure  
that the motors are running...

# Think Aloud Usability Test

- <https://www.youtube.com/watch?v=SrU5pLGs3JQ>
- <https://www.youtube.com/watch?v=-h8hUtwkMCE>

# Predictive Evaluation

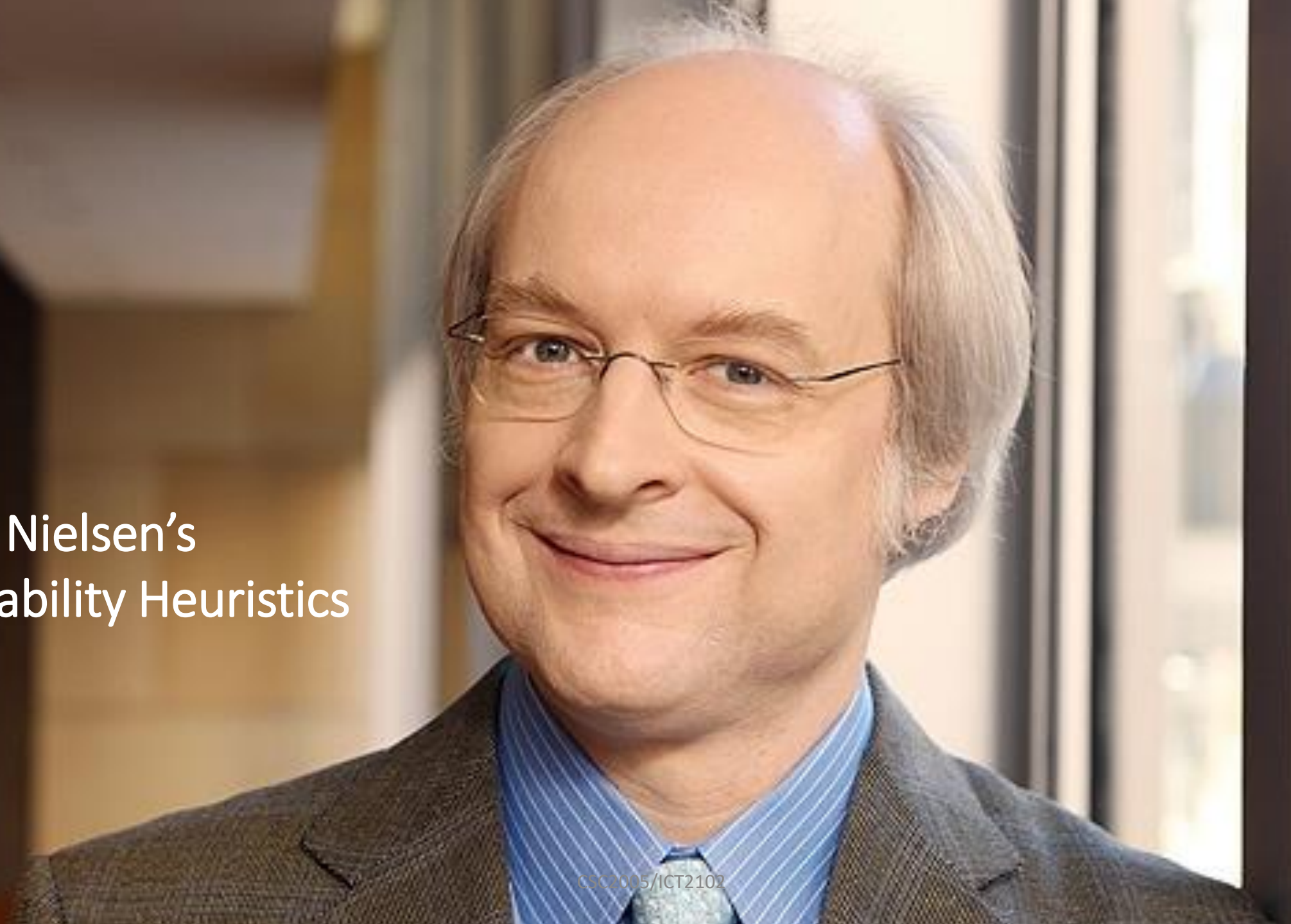
- Conducted with expert reviewers in the field
- HCI experts (not real users) interact with the system, find potential problems, and give prescriptive feedback
- Works best if they:
  - Haven't used earlier prototype
  - Are familiar with domain or task
  - Understand user perspectives



# Predictive Evaluation Methods

- Heuristic Evaluation
- Cognitive Walkthrough
- Many other methods

# Jakob Nielsen's 10 Usability Heuristics



# Heuristic Evaluation

- Developed by Jakob Nielsen in the early 1990s
- Heuristics are developed based on broad “rules of thumb”
- Small set (3-5) expert usability evaluators assess system based on simple and general heuristics (principles or rules of thumb)
  - 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 sketches
- <http://www.nngroup.com/articles/ten-usability-heuristics/>

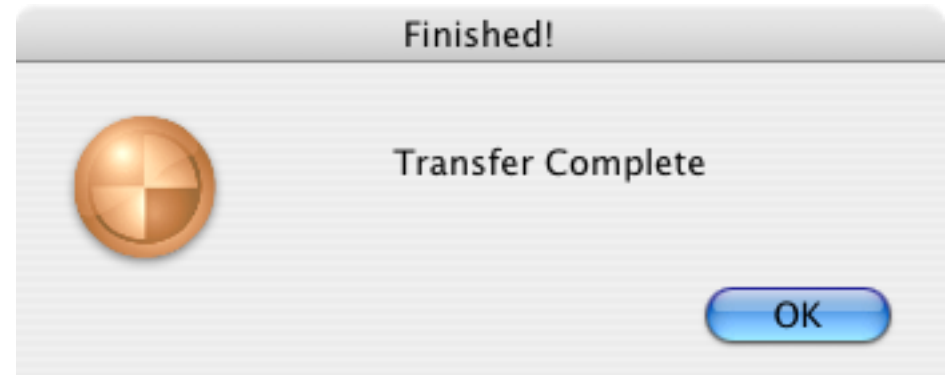
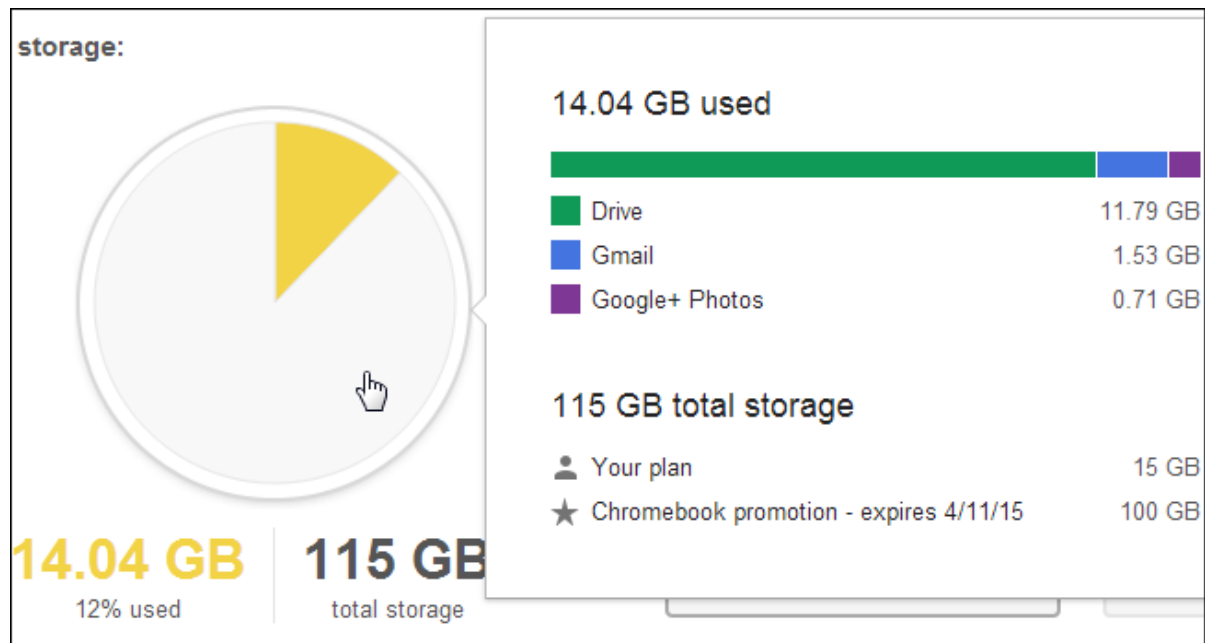
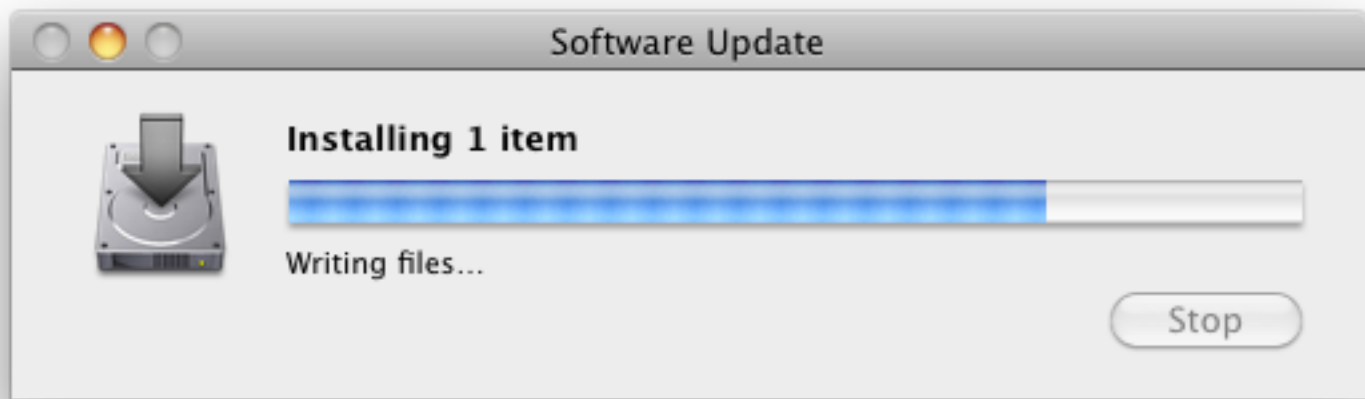
# Ten Usability Heuristics

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

# 1/10 Visibility of system status

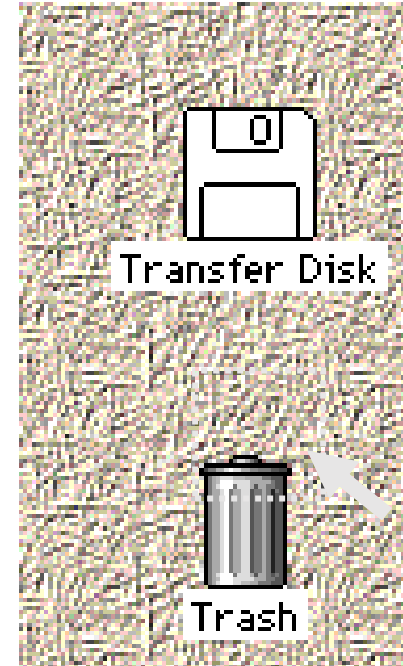
- Feedback of system's current state
- Page is loading, error loading page, something saved or deleted
- Show completion and what to do next
- Can the (internal) system state be seen or reflected to the user?

<https://www.nngroup.com/articles/visibility-system-status/>



## 2/10 Match between system and the real world

- Use familiar metaphors and language
- Follow real world conventions
- Speak the user's language
- Familiar categories, language and choices
- Familiar components and widgets



Old Mac Desktop: Dragging disk to trash should delete it, not *eject* it



CART



LIBRARY



PLATFORM



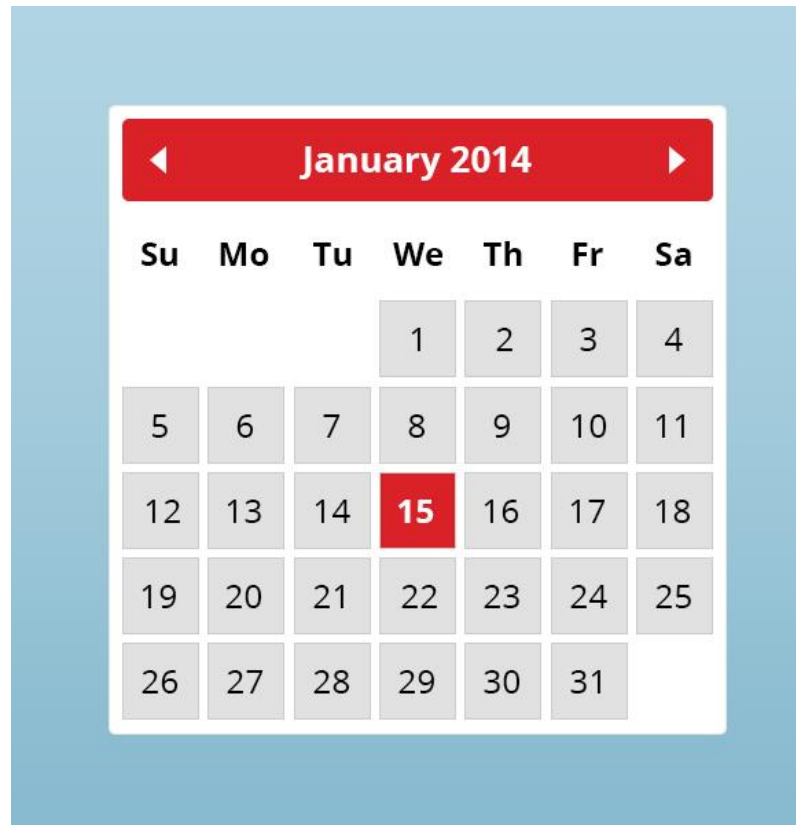
DELETE



State

Select a

- Select a State
- Alabama
- Alaska
- Arizona
- Arkansas**
- California
- Colorado
- Connecticut
- D.C.
- Delaware
- Florida
- Georgia
- Hawaii
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Louisiana



# 3/10 User control and freedom

- Freedom to undo, explore and change choices
- “Exits” for mistaken choices
- Exploration of data or options
- Multiple methods of doing the same thing
- Do tasks the way they like

MARCH						
SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
23	24	25	26	27	28	1
2	3	4	5	6	7	8
9	10	FROM \$194 ONE-WAY/PP	FROM \$194 ONE-WAY/PP	FROM \$194 ONE-WAY/PP	FROM \$139 ONE-WAY/PP	FROM \$139 ONE-WAY/PP
16	FROM \$139 ONE-WAY/PP	FROM \$49 ONE-WAY/PP	FROM \$49 ONE-WAY/PP	FROM \$84 ONE-WAY/PP	FROM \$139 ONE-WAY/PP	FROM \$104 ONE-WAY/PP
23	FROM \$104 ONE-WAY/PP	FROM \$49 ONE-WAY/PP	FROM \$49 ONE-WAY/PP	FROM \$49 ONE-WAY/PP	FROM \$89 ONE-WAY/PP	FROM \$79 ONE-WAY/PP
30	FROM \$69 ONE-WAY/PP					

DECEMBER						
SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
25	26	27	28	29	30	FROM 13,900pts +\$5.00
2	FROM 13,900pts +\$5.00 0 seats left	FROM 13,900pts +\$5.00	FROM 13,900pts +\$5.00	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50
9	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50	FROM 9,600pts +\$2.50
16	FROM 11,500pts +\$2.50	FROM 11,500pts +\$2.50	FROM 15,500pts +\$2.50	FROM 14,000pts +\$2.50 4 seats left	FROM 30,000pts +\$2.50	FROM 48,000pts +\$5.00
23	FROM 30,000pts +\$2.50	FROM 18,500pts +\$2.50 2 seats left	FROM 18,500pts +\$5.00	FROM 30,000pts +\$2.50	FROM 18,500pts +\$2.50 1 seat left	FROM 11,500pts +\$2.50
						FROM 14,000pts +\$2.50

Compare  
fare  
options

Prices displayed are one-way per person, excluding taxes and fees. Additional [baggage fees](#) may apply.  
Did you know you can click on a flight number to see its on-time performance?

Departs ▾	Arrives ▾	Duration ▾	BLUE ▾	BLUE PLUS ▾	BLUE FLEX ▾
11:00 AM FLL ✈ 387** A320	12:12 PM SNU <a href="#">View seats</a>	1h 12m <a href="#">Amenities</a>	 3,500 pts + \$5.60 taxes/fees	 4,600 pts + \$5.60 taxes/fees	 10,900 pts + \$5.60 taxes/fees

## Returning flights

Santa Clara, Cuba (SNU) to Fort Lauderdale, FL (FLL)

\*\*This flight is subject to government approval

← Prev	Fri Sep 9 3,500 pts	Sat Sep 10 N/A	Sun Sep 11 N/A	<b>Mon Sep 12 3,500 pts</b>	Tue Sep 13 N/A	Wed Sep 14 3,500 pts	Thu Sep 15 N/A	→ Next
-----------	------------------------	-------------------	-------------------	---------------------------------	-------------------	-------------------------	-------------------	-----------

Compare  
fare  
options

Prices displayed are one-way per person, excluding taxes and fees. Additional [baggage fees](#) may apply.  
Did you know you can click on a flight number to see its on-time performance?

Departs ▾	Arrives ▾	Duration ▾	BLUE ▾	BLUE PLUS ▾	BLUE FLEX ▾
1:15 PM SNU ✈ 388** A320	2:35 PM FLL <a href="#">View seats</a>	1h 20m <a href="#">Amenities</a>	 3,500 pts + \$41.46 taxes/fees	 4,600 pts + \$41.46 taxes/fees	 10,900 pts + \$41.46 taxes/fees

### FILTER RESULTS

☐ Nonstop flights

### SELECT FARE

☒ TrueBlue points

☐ Dollars

### ITINERARY

Estimated total 1 traveler

**7,000 PTS**  
**+ \$47.06 USD**

**ZERO IN  
ON A \$0  
ANNUAL FEE.\***



jetBlueCard

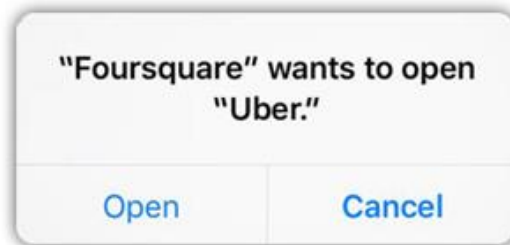
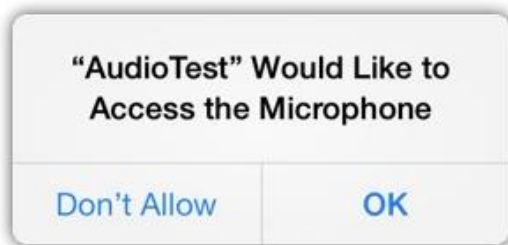
[Learn More](#)

\*See Terms and Conditions for details.

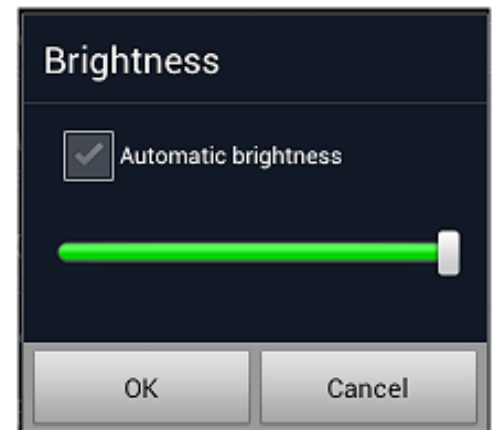
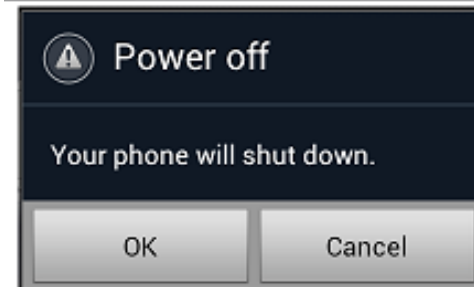
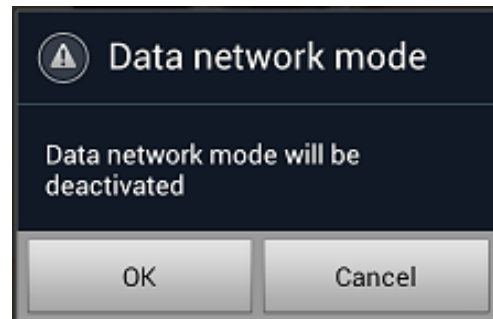
# 4/10 Consistency and standards

- Consistent layout
- Standard locations and positions
  - Same language, placement etc. everywhere
  - Follow platform conventions
- Same terminology is used
- Same structure on all webpages
  - E.g. submit button at the bottom after filling out information



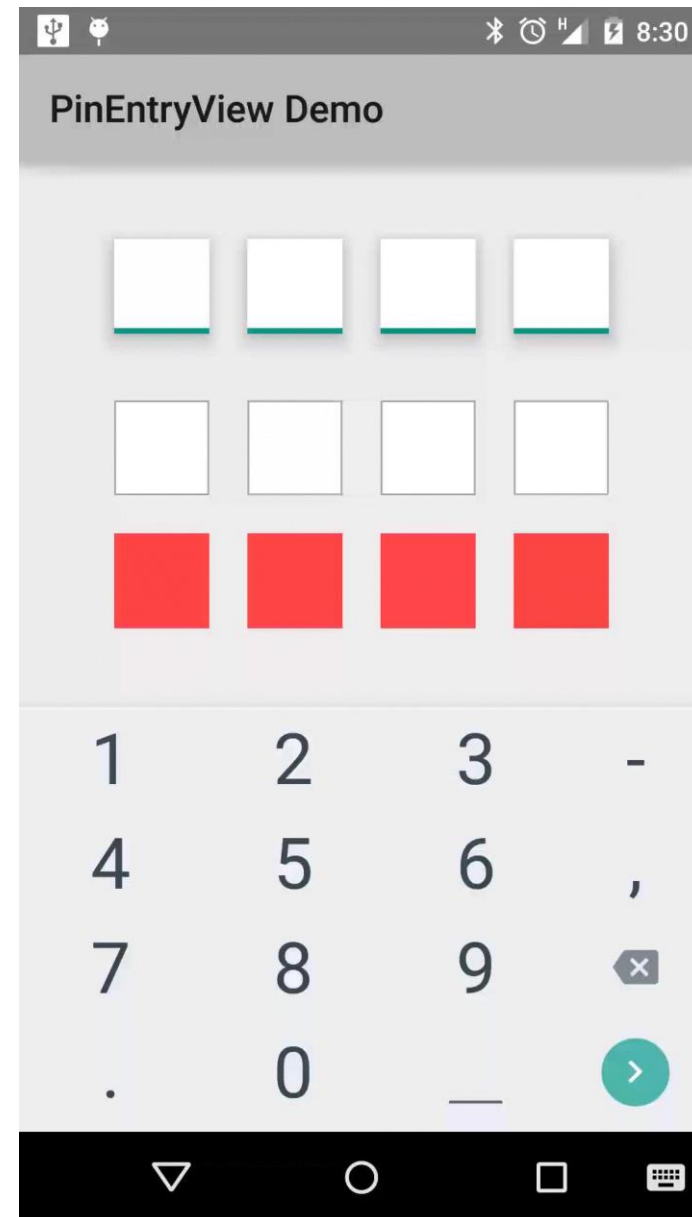
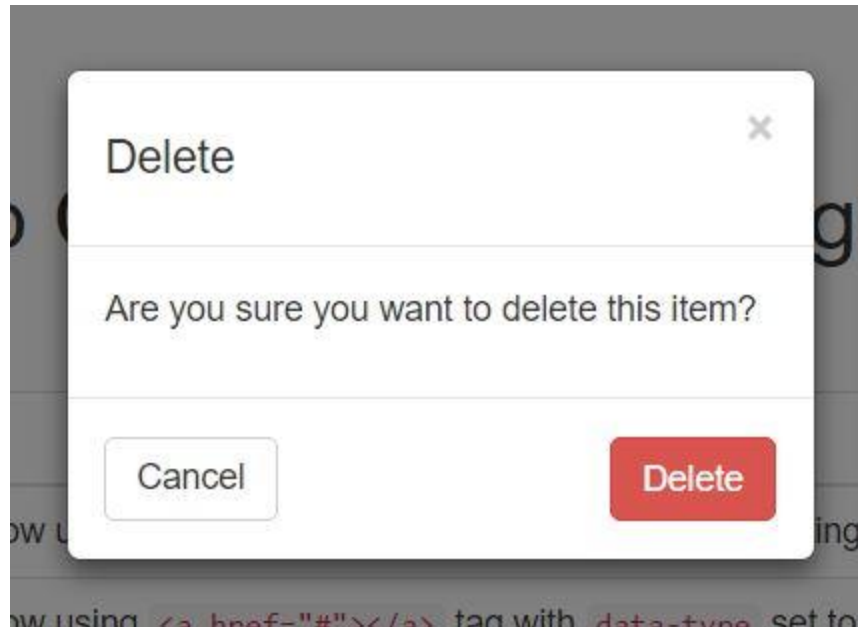


iOS uses action priority rather than static button order

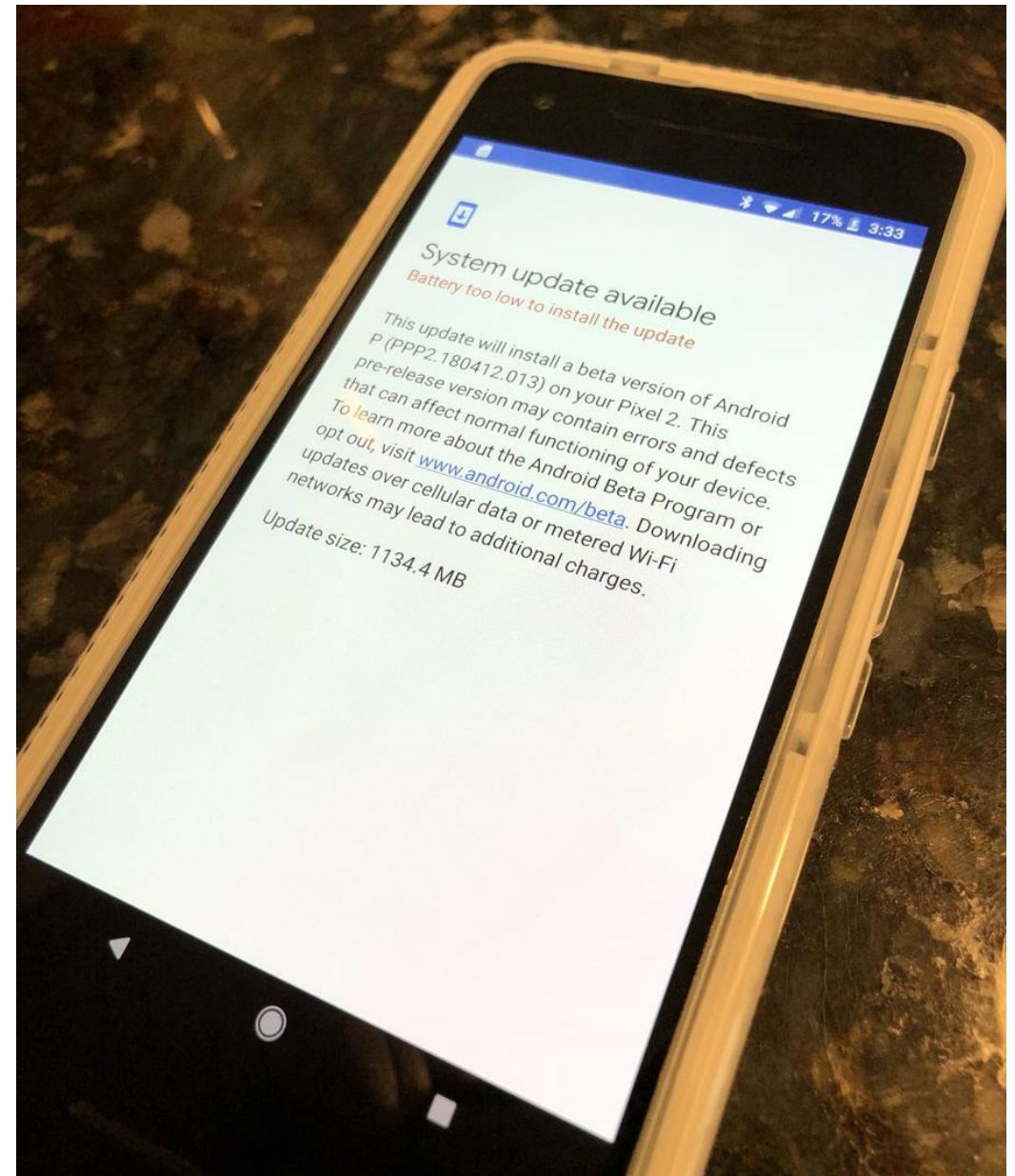
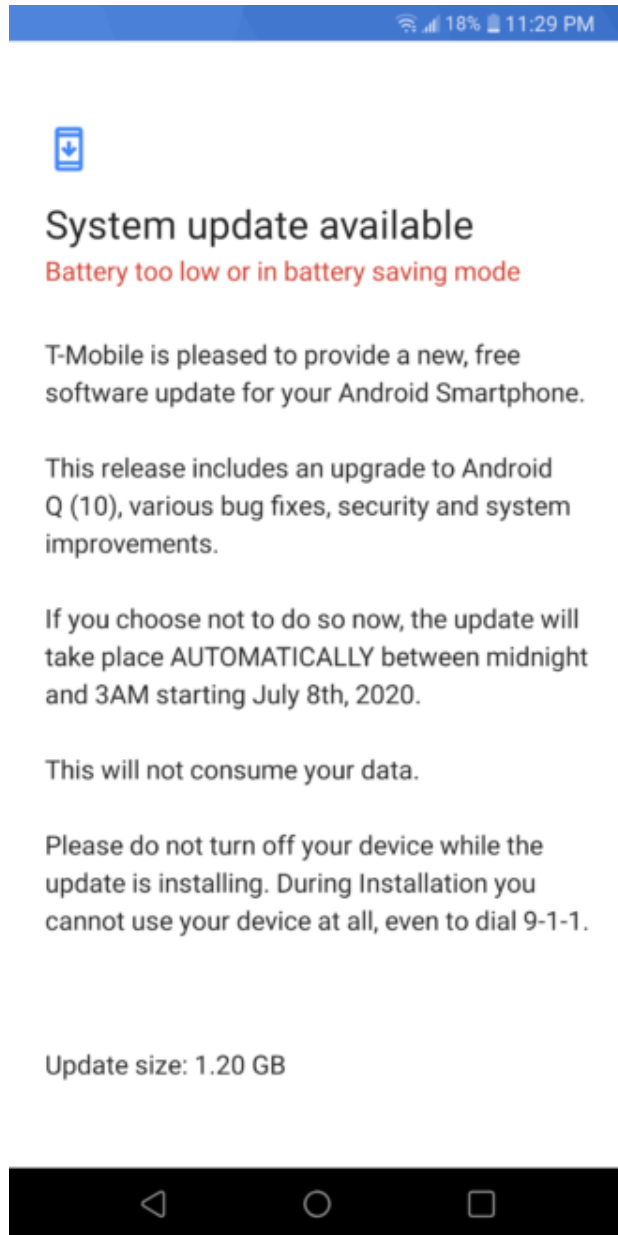


# 5/10 Error prevention

- Prevent data loss
  - e.g. Confirmation dialog before deleting an item
- Prevent bad input
  - e.g. restrict to numerals if asked to key in mobile number
- Prevent misinterpretation of the message
  - e.g. User entered text in the field when entering their mobile number. The system also displays errors about compulsory fields not completed, but the problem is with the mobile number input



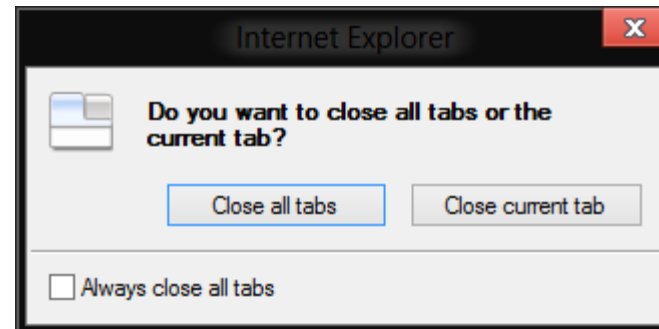


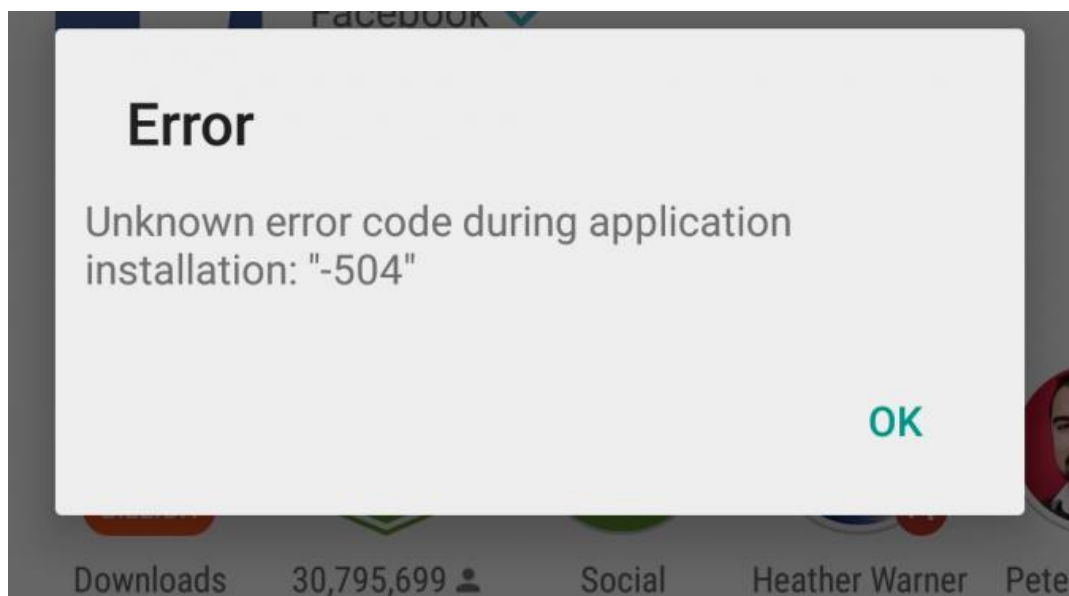


# 6/10 Recognition rather than recall

- Make objects, actions, options, & directions visible/easily retrievable
- Show users previews e.g. images, results
- Do not use codes
- Icons to click rather than type the text

```
% rm -rf *  
%
```





# 7/10 Flexibility & efficiency of use

- Flexible shortcuts
  - accelerators for experts (e.g., gestures, kb shortcuts)
  - allow users to tailor frequent actions (e.g., macros)
- Defaults with options
- Recommendations and relevant items

Google

Mail

COMPOSE

Inbox

Starred

Important

Sent Mail

Drafts

All Mail

Spam (3)

Trash

Keyboard shortcuts

Keyboard shortcuts are enabled. [Disable](#) [More](#)

**Jumping**

- g** then **i** : Go to Inbox
- g** then **s** : Go to Starred conversations
- g** then **t** : Go to Sent messages
- g** then **d** : Go to Drafts
- g** then **a** : Go to All mail
- g** then **c** : Go to Contacts
- g** then **k** : Go to Tasks
- g** then **l** : Go to Label

Using 1.5 GB of your 10.1 GB

**Threadlist selection**

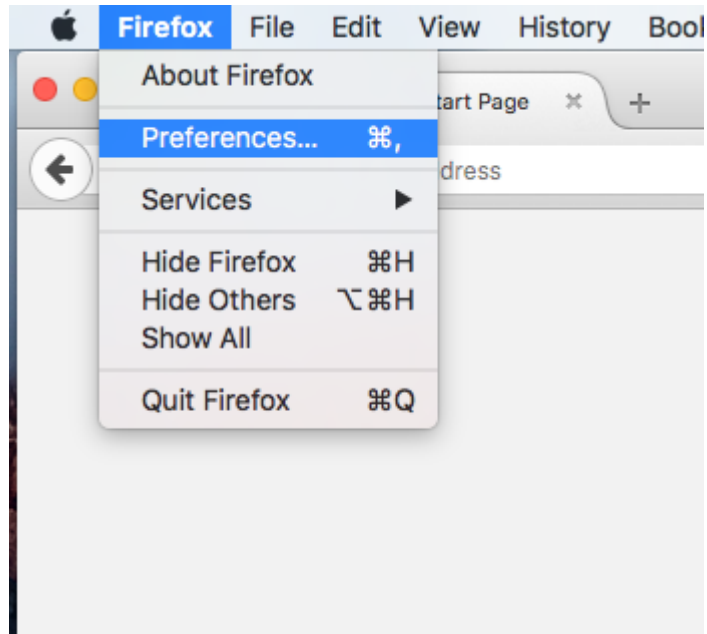
- \*** then **a** : Select all conversations

No new mail!

**Actions**

- ,** : Move focus to toolbar
- x** : Select conversation
- s** : Rotate superstar
- y** : Remove label
- e** : Archive
- m** : Ignore conversation
- !** : Report as spam
- #** : Move to trash
- r** : Reply
- <Shift> + r** : Reply in a new window

©2012 Google - [Terms of Service](#) - [Privacy Policy](#) - [Program Policies](#)

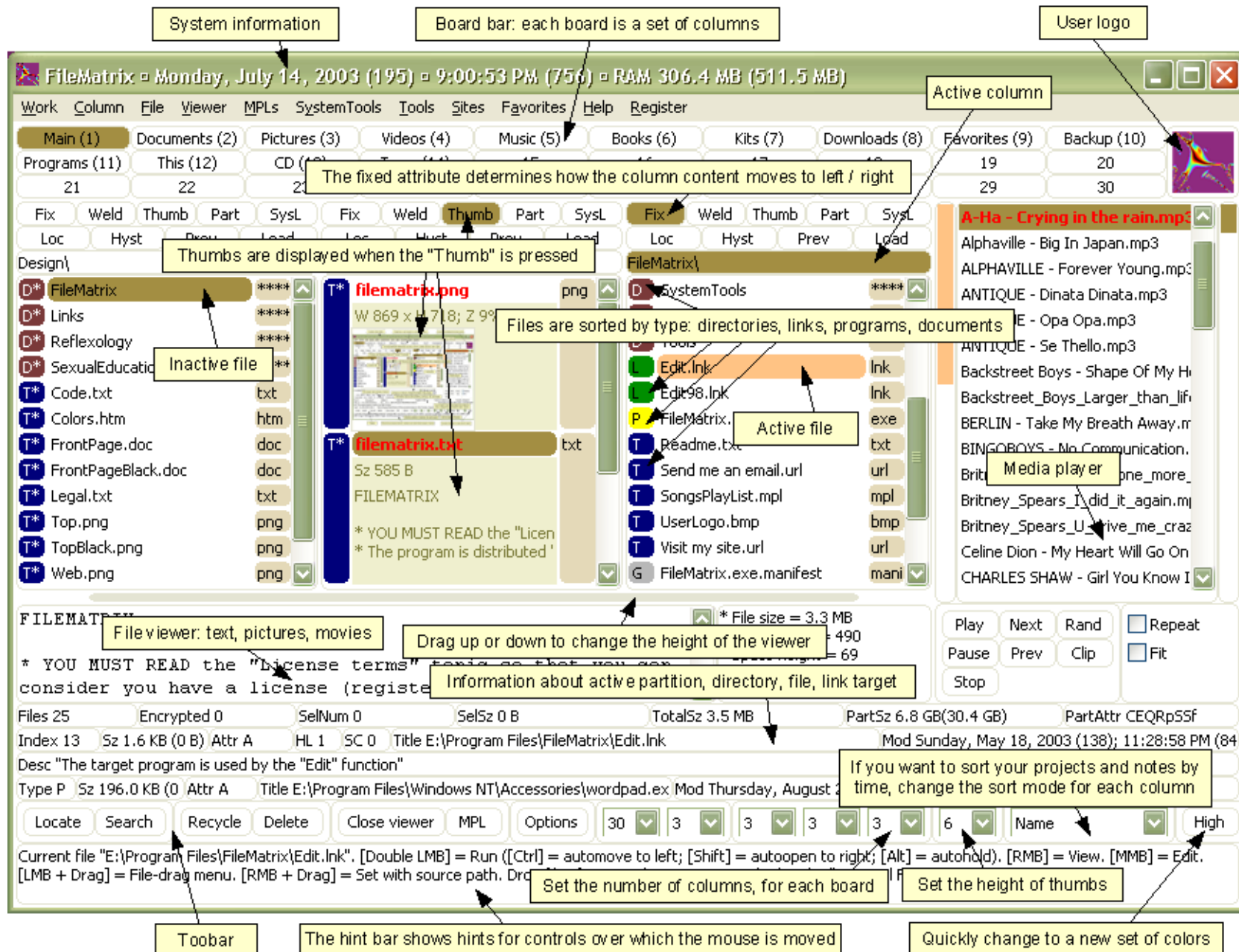


amazon.com

### More Items to Consider

You viewed	Customers who viewed this also viewed
<div><p><b>The Appeal</b> John Grisham Paperback <del>\$14.00</del> <b>\$11.20</b></p></div>	<div><p><b>The Innocent Man</b> John Grisham Mass Market Paperback <b>\$7.99</b></p></div> <div><p><b>The Associate: A Novel</b> John Grisham Mass Market Paperback <b>\$9.99</b></p></div> <div><p><b>Ford County: Stories</b> John Grisham Paperback <del>\$15.00</del> <b>\$8.19</b></p></div>





Note: I live inside this website Monday to Friday 9am-6pm, to give you the very best service and make your experience a happy one! - I am Ling, accept no substitutes

**LINGSCARS.com**

UK CONTRACT HIRE CARS FROM LING VALENTINE

I prefer email to phone ☺  
 sales@LINGSCARS.com  
 T 0191 460 9444  
 F 0870 486 1130  
 World HQ, Vance Bus Pk  
 Gateshead, NE11 9NE

CAR and VAN PRICES - have a look... INFORMATION - have a think... ACTION - have a go!

**FULL BUSINESS CAR PAGE** **FULL PRIVATE CAR PAGE** **VANS AND PICKUPS** **GET P/X PRICE** **Step by step guide** **HELLO! I am Ling** **Ling's tips & hints!** **1153 letters! Big page!** **175 PRESS Items!** **GREEN car info** **Get unlisted car QUOTE!** **ORDER a car now!**

commended as **BEST non-franchised WEBSITE 2007** **EXTRA CHEAP** **Chinese Car Websites!** **CREDIT CHECK? DIY, free, HERE** **my BLOG!** **RSS/XML feed-Wah!** **commended as BEST non-franchised 2008 ENVIRONMENTALIST** **FREE CASH blatant bribery from LING! CLICK!**

**DRAGONS' DEN SUCCESS READ!** **NORTH-EAST ENTREPRENEUR OF THE YEAR Winner of the Woman in Retail Award: LING VALENTINE** **WEAR & TEAR GUIDE HERE** **Customer AZ Play with my customers! ...interactive map toy Ling-oogle Maps** **My Top 20 cars! BEST-SELLING CARS!** **VIZ OFFICIAL Business Ambassador!**

Office hours 11to1 Ling live chat: **Live Chat is Closed** **CHEAP car insurance:** **Click HERE** **MY BUS HERE**

**WIN! Ling's RACING tips!** **Current Winnings: £-74.35**

**Less fuel = Less CO2 NEW CARS SAVE FUEL!** **Shell** **weekly fuel costs displayed for EVERY SINGLE CAR** **CLICK HERE**

**NEW Ford Fusion 1.4 TDCi (67bhp) Style Hatchback 5dr 1399cc** **£169.99/month (plus VAT)** **World Famous CHOP-GEAR TV**

**CAR-U-LIKE-ATOR** **PetroChina 中国石化** **2008 LINGSCARS.com RALLYCROSS OPEN CHAMPIONSHIP**

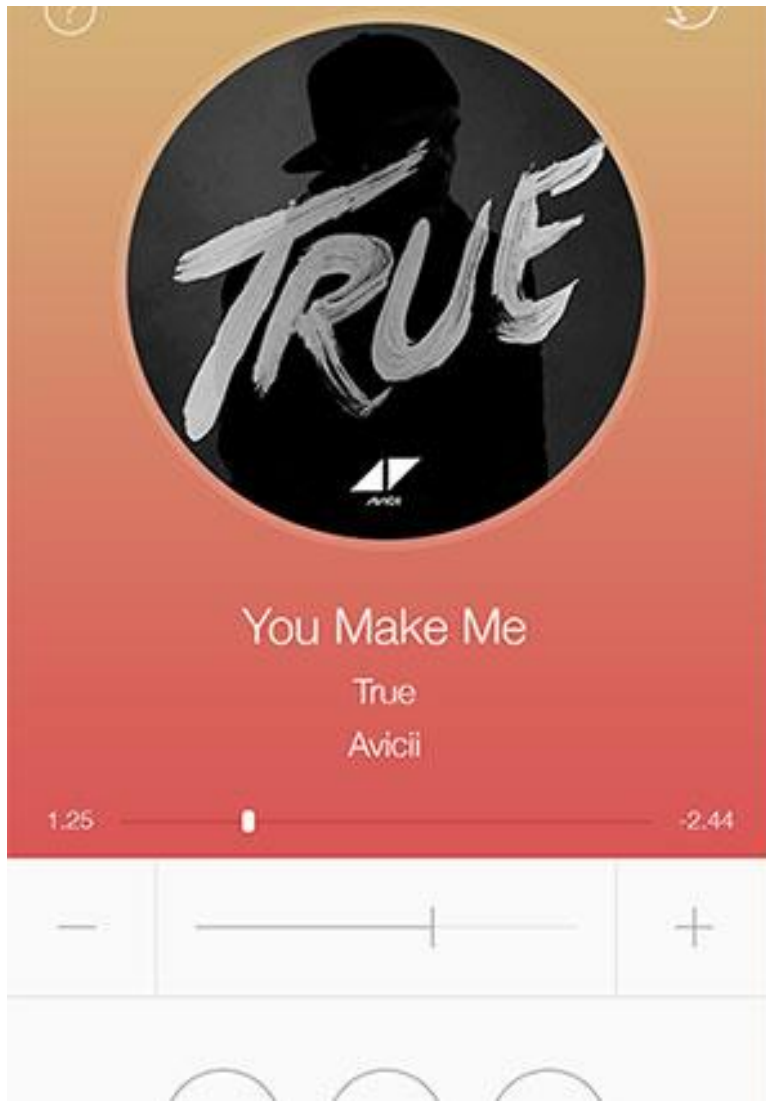
**PERSONAL CAR PRICE LIST** **Automatic email updates KEEP YOUR EYE ON A CAR click here!** **BE THE 1st TO KNOW knowledge is power - BACON & LING**

**CALLING ALL VISITORS!** **I am human being, not robot! Office hours: Mon-Fri 9 to 6 ...after 6pm I making bean curd!** **Ling Tube Broadcast Herself™** **...enjoy my website...**



# 8/10 Aesthetic & minimalist design

- Clean and uncluttered interface
- Eliminate redundant information, only leave what has to be there
- No irrelevant information in dialogs





Welcome to  
**Mail App**

Enter your name



Password

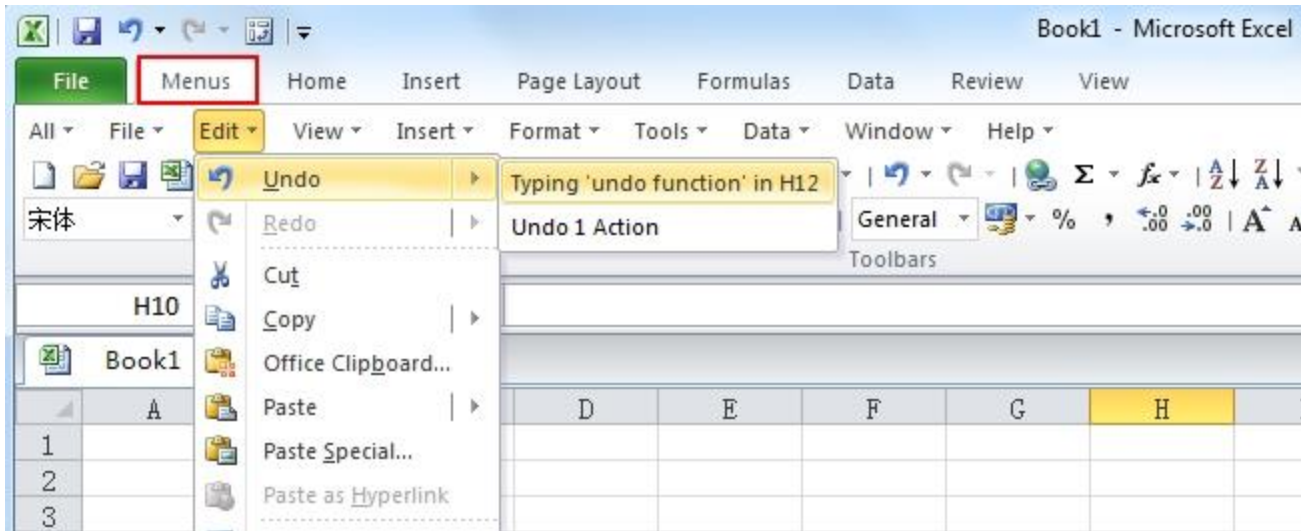


Login

[Lost your password?](#)

# 9/10 Recognize, diagnose, & recover from errors

- Clearly state the problem, provide a solution and how to address the problem
  - E.g. Created a password with 8 characters, but system requires at least 1 special character and 1 numeral. Error message should state the requirements, so the user can fix the error
  - E.g. Did not fill all the fields while registering for an account. Error message should highlight the fields which have not been filled



←
Application
⋮

### Personal info

First name

First name is required

Connors

01/01/90

MM/DD/YYYY

Social Security

### - ## - ####

### Residence address

2000 Main Street, Apartment J

San Francisco

CA

94112

United States

# 10/10 Help & documentation

- Provide more information and clarify, show steps and what to do next
  - easy to search
  - focused on the user's task
  - list concrete steps to carry out
- E.g. press (?) beside a field, popup will tell you what the field is about, and what you should enter
- E.g. Help text on social networking websites for users to understand the consequences of certain things – upload photos, change privacy settings, who can see your posts etc.

#### Phone Details

IMEI Code:  ?

PAC Code:  ?

#### Phone Details

IMEI Code:  ?

#### Explanation of IMEI Code

The International Mobile Equipment Identity (IMEI) number is a unique 15-digit code used to identify an individual GSM mobile telephone. The number can be found on most mobiles by typing in \*#06#. If this combination doesn't work on your mobile phone, please call our support centre on +44 (0) 1252 xxxx xxx.

[Back to IMEI input field.](#)

PAC Code:  ?

# Phases of Heuristic Evaluation

1. Pre-evaluation training
  - Give evaluators needed domain knowledge and information on the scenario
2. Evaluation
  - Individuals evaluate and then aggregate results
3. Severity rating
  - Determine how severe each problem is (priority). Can do first individually and then as a group
4. Debriefing
  - Review with design team



# How to perform heuristic evaluation

- At least two passes for each evaluator
  - 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

# How to perform heuristic evaluation

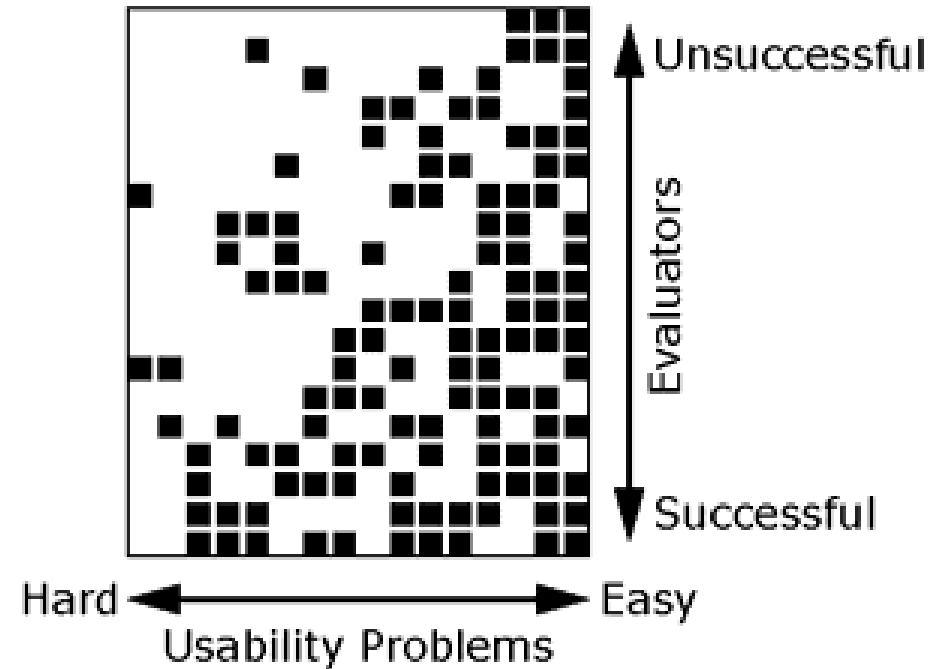
- Why separate listings for each violation?
  - risk of repeating problematic aspect
  - may not be possible to fix all problems
- Where problems may be found
  - single location in UI
  - two or more locations that need to be compared
  - problem with overall structure of UI
  - something that is missing
    - common problem with paper prototypes
    - note: sometimes features are implied by design docs and just haven't been "implemented" – relax on those

# Evaluators' Process

- Step through design several times
  - Examine details, flow, and architecture
  - Consult list of usability principles
  - ..and anything else that comes to mind
- Evaluate against heuristics
  - Nielsen's "heuristics"
  - Category-specific heuristics
  - e.g., design goals, competitive analysis, existing designs
- Use violations to redesign/fix problems

# Multiple Evaluators

- No evaluator finds everything
- Some find more than others
- Single evaluator achieves poor results
  - Only finds 35% of usability problems
- 5 evaluators find ~ 75% of problems



# Severity Rating

- Independently estimate after review
- Allocate resources to fix problems
- Estimate need for more usability efforts
- Severity combines:
  - Frequency: how common?
  - Impact: how hard to overcome?
  - Persistence: how often to overcome?

# Severity Ratings

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

# Severity Ratings Example

## 1. [H1-4 Consistency] [Severity 3][Fix 0]

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.

# Debriefing

- Conduct with evaluators, observers, and development team members
- Discuss general characteristics of UI
- Suggest potential improvements to address major usability problems
- Development team rates effort to fix
- Brainstorm solutions



# Cost-effectiveness of Heuristic Evaluations

- In one case: benefit-cost ratio of 48
  - estimated benefit \$500,000; cost \$10,500
  - value of each problem ~\$15K
- How to calculate this value?
  - in-house -> productivity
  - open market -> sales
- Severe problems found more often

# Heuristics vs. User Testing

- Heuristic evaluation often faster
  - 1-2 hours for each evaluator
- Heuristic evaluation results come pre-interpreted
- User testing is more accurate (by definition)
  - Takes into account actual users and tasks
  - Heuristic evaluation may miss problems & find “false positives”
- Valuable to alternate methods
  - Find different problems
  - Don't waste participants

# Shneiderman's 8 Golden Rules

1. Consistency
2. Shortcuts
3. Feedback
4. Dialog closure
5. Simple error handling
6. Reversible actions
7. Put user in control
8. Reduce short-term memory load

<https://www.cs.umd.edu/users/ben/goldenrules.html>

# Cognitive Walkthrough

- Assess learnability and usability through simulation of the way users explore and become familiar with interactive system
- Essentially a usability “thought experiment”
- Like code walkthrough in software engineering
- Polson, Lewis, et al. at UC Boulder (2002)

# Cognitive Walkthrough Process

- Construct carefully designed tasks from system spec or screen mock-up
- Walk through (cognitive & operational) activities required to go from one screen to another
- Review actions needed for task, attempt to predict how users would behave and what problems they'll encounter

# Cognitive Walkthrough Requirements

- Description of users and their backgrounds
- Description of task user is to perform
- Complete list of the actions required to complete task
- Prototype or description of system

# Cognitive Walkthrough Assumptions

- User has a rough plan (not always true)
- User explores system, looking for actions to contribute to performance
- User selects action that seems best for reaching desired goal
- User interprets response and assesses whether progress has been made toward completing task

# Cognitive Walkthrough Methodology

- Step through action sequence
  - Action 1
  - Response A, B, ..
  - Action 2
  - Response A
  - ...
- For each one, ask **four questions** and try to construct a believability story



# Cognitive Walkthrough 4 Questions

1. Will users be trying to produce whatever effect the given action has?
2. Will users be able to notice that the correct action is available?
3. Once found, will they know it's the right action for the desired effect?
4. Will users understand feedback after the action?

# Question 1

- Will user be trying to produce effect?
- Typical supporting evidence
  - It is part of their original task
  - They have experience using the system
  - The system tells them to do it
- No evidence?
  - Construct a failure scenario
  - Explain, back up opinion

# Question 2

- Will user notice the action is available?
- Typical supporting evidence
  - Experience
  - Visible device, such as a button
  - Perceivable representation of an action such as a menu item

# Question 3

- Will user know it's the right action for the desired effect?
- Typical supporting evidence
  - Experience
  - Interface provides a visual item (such as prompt) to connect action to result effect
  - All other actions look wrong

# Question 4

- Will user understand the feedback?
- Typical supporting evidence:
  - Experience
  - Recognize a connection between a system response and what user was trying to do

# Cognitive Walkthrough Examples

- <https://www.youtube.com/watch?v=CeWAbGU5cSw>
- <https://www.youtube.com/watch?v=Edqjao4mmxM>

# DECIDE: An Evaluation Framework

**D**etermine the aims and goals

**E**xplore the questions you will ask

**C**hoose the evaluation approach/method

**I**dentify practical issues

**D**ecide how to deal with ethical issues

**E**valuate, analyze, interpret and present the data collected

# Determine the aims and goals

- What are the high level objectives of the evaluation ?
- Who wants it and why ?
- The goals influence the paradigm for the study
- Some examples of goals:
  - Identify the best metaphor to base the design
  - Check to ensure that the final interface is consistent
  - Investigate how technology affects working practices



# Explore the questions

- All evaluations need goals & questions to guide them so time is not wasted on ill defined studies
- Questions are based on the goals of the study
- Example: the goal of finding out how many customers prefer to purchase paper airline tickets rather than e-tickets can be broken down into sub-questions:
  - What are the customers' attitudes to these new tickets?
  - Are they concerned about security?
  - Is the interface for obtaining them poor?

# Choose the evaluation paradigm & techniques

- An evaluation paradigm is an approach that is influenced by particular theories and philosophies
- The evaluation paradigm strongly influences the techniques used, and how the data is analyzed and presented
- E.g. field studies do not involve testing and modeling

# Identify practical issues

For example, how to:

- Select users
  - where to find subjects?
  - how to contact them?
  - How many subjects do I need?
- Stay on budget
- Staying on schedule
- Find evaluators
- Select equipment

# Decide on ethical issues

- Develop an informed consent form
- Participants have a right to:
  - Know the goals of the study
  - What will happen with the findings
  - Privacy of personal information
  - Not to be quoted without their agreement
  - Leave when they wish
  - Be treated politely
- Ethics board (aka Institutional Review Board, IRB)
  - Need permission from ethics committee
  - Review documentation, approve

# Evaluate, interpret and present data

- How data is analyzed and presented depends on the paradigm and techniques used
- The following also needs to be considered:
  - Reliability: can the study be replicated?
  - Validity: is it measuring what you thought?
  - Biases: is the process creating biases?
  - Scope: can the findings be generalized?
  - Ecological validity: is the environment of the study influencing it?  
e.g. Hawthorn effect

# Summary (1)

- There are many methods for conducting an evaluation, dependent on the stage in the design process and goals and objectives
- Empirical evaluation is verifiable or provable by means of observation or experiment (based on empirical evidence)
- Predictive evaluation methods are conducted with expert reviewers in the field
- A heuristic evaluation makes use of 10 usability “rules of thumb”
- A cognitive walkthrough is a usability evaluation method in which one or more evaluators work through a series of tasks and ask a set of questions from the perspective of the user

# Summary (2)

- Many issues to consider before conducting an evaluation study
- Includes the goals of the study, the approaches and methods to use, practical issues, ethical issues, and how the data will be collected, analyzed and presented.
- The DECIDE framework provides a useful checklist for planning an evaluation study

# References

- *Usability Engineering*, by Nielsen, 1994
- <http://www.nngroup.com/articles/>
- <https://www.cs.umd.edu/users/ben/goldenrules.html>



Write a task list, and do a “think aloud” of one of the vending machines on campus

Conduct a heuristic evaluation  
of one of the vending machines  
on campus