Usability Evaluation

Scott Klemmer, Bill Verplank
ta: Kevin Collins
and Nundu JanakiRam

22 February 2007

Today: Meet together

1:15 - 1:25 Announcements

1:25 - 2:00 Usability Evaluation

2:00 - 2:05 *Split into sections*

2:05 - 3:05 Moving Forward with P4

Web survey results

- 10 responses…
- Excellent (2)
- Very good (5)
- fair (2)
- poor (1)

Tutorials

- On the whole, very well received
- "They're all fun and interesting, I especially enjoyed the tutorials with the mobile phones."
- "this class gave me a chance to do things that I would have probably never done otherwise, such as soldering together my own light switch - my dad couldn't believe I did something like that!"
- But…
  - "If the PRL safety training could be scheduled later on in the quarter"

Content

- Strengths:
  - "Learning to design for real people"
  - "The exercises in class"
  - "The fact that it hasn't seemed like a typical course (i.e. with lectures, weekly sections, midterm & final)"
  - "Brainstorming, drawing, and storyboarding techniques."
- "Can we see more examples of interfaces that have gone wrong?"
- Q: Are these methods used in industry? A:
  - Many (e.g., design notebooks, storyboards, discount evaluation methods) are very common

Organization

- "I love how the course is very well-organized"
- "A little disorganized sometimes, in that I don't know what I'm supposed to be doing"
- Classtime usage
  - "The excercises in class and the class tutorials have been engaging, unique and very interesting."
  - "The in-room activities don't seem to be very enlightening."
- "overlap with cs147 material"

Hand-ins

- The challenges of a multidisciplinary course…
- "P2 was extremely open-ended" & "Clearer deliverables" v. "This class needs to be run more like a studio art class, where the work is entirely student-directed"
- "More clear guidelines about when and what to turn in."

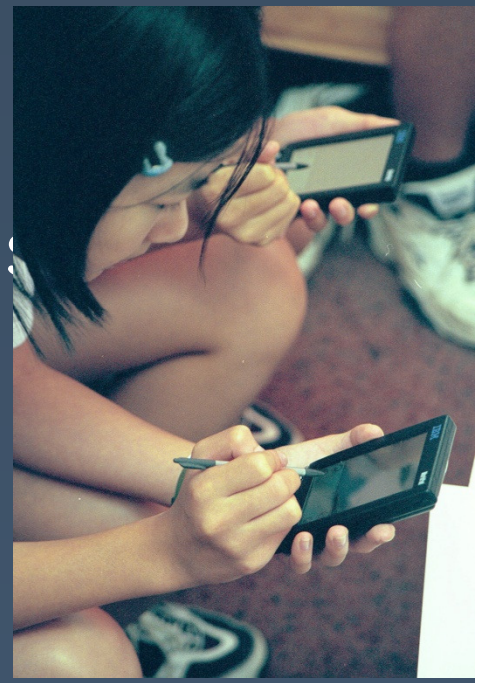
Usability Evaluation

Scott Klemmer, Bill Verplank
TA: Kevin Collins
and Nundu JanakiRam

22 February 2007

Evaluation Outline

- Lab-based user testing
- Discount usability
- Remote / Online user tes
- Action analysis
- Heuristic evaluation

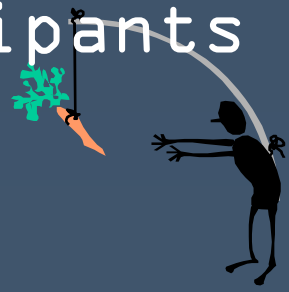


Why do User Testing?

- Can't tell how good UI is un
  - people use it!
- Other methods are based on evaluators who
  - may know too much
  - may not know enough (about tasks, etc.)
- Hard to predict what real users will do

Choosing Participants

- Representative of target u
  - job-specific vocab / knowled
  - tasks
- Approximate if needed
  - system intended for doctors
    - get medical students
  - system intended for engineer
    - get engineering students
- Use incentives to get participants

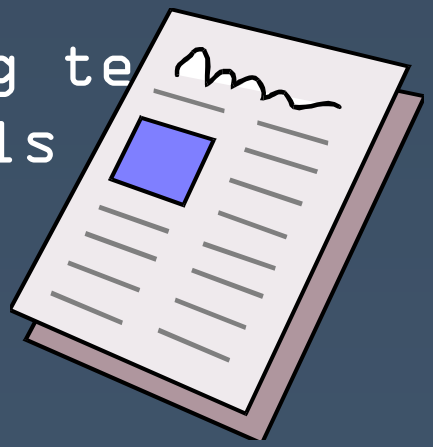


Ethical Considerations

- Sometimes tests can be distressing
  - users have left in tears
- You have a responsibility to alleviate
  - make voluntary with informed consent
  - avoid pressure to participate
  - let them know they can stop at any time
  - stress that you are testing the system, not them
  - make collected data as anonymous as possible
- Often must get human subjects approval

User Test Proposal

- A report that contains
  - objective
  - description of system being te
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures

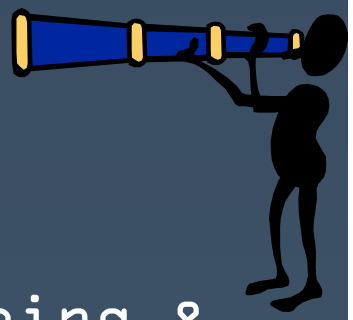


- Get approved & then reuse for final report
- Seems tedious, but writing this will help “debug” your test

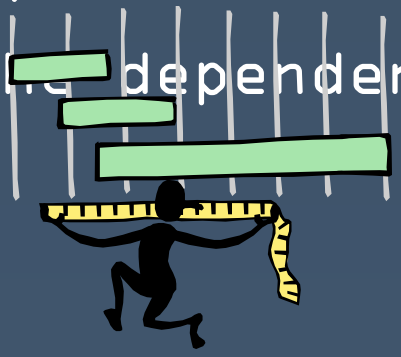
Selecting Tasks

- Should reflect what real tasks will be like
- Tasks from analysis & design can be used
  - may need to shorten if
    - they take too long
    - require background that test user won't have
- Try not to train unless that will happen in real deployment
- Avoid bending tasks in direction of what your design best supports

Deciding on Data to Collect



- Two types of data
  - process data
    - observations of what users are doing & thinking
  - bottom-line data
    - summary of what happened (time, errors, success)
    - i.e., the dependent variables



Which Type of Data to Collect?

- Focus on process data first
  - gives good overview of where problems are
- Bottom-line doesn't tell you where to fix
  - just says: "too slow", "too many errors", etc.
- Hard to get reliable bottom-line results
  - need many users for statistical significance

The “Thinking Aloud” Method

- Need to know what users are thinking, not just what they are doing
- Ask users to talk while performing tasks
  - tell us what they are thinking
  - tell us what they are trying to do
  - tell us questions that arise as they work
  - tell us things they read
- Make a recording or take good notes
  - make sure you can tell what they were doing

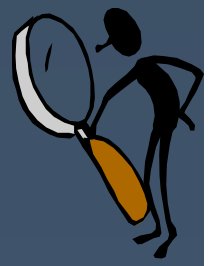
Thinking Aloud (cont.)

- Prompt the user to keep talking
  - "tell me what you are thinking"
- Only help on things you have pre-decided
  - keep track of anything you do give help on
- Recording
  - use a digital watch/clock
  - take notes, plus if possible
    - record audio & video (or even event logs)

Video of a Test Session

Using the Test Results

- Summarize the data
  - make a list of all critical incidents (CI)
    - positive & negative
  - include references back to original data
  - try to judge why each difficulty occurred
- What does data tell you?
  - UI work the way you thought it would?
    - users take approaches you expected?



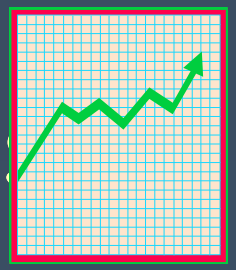
Using the Results (cont.)

- Update task analysis & rethink design
  - rate severity & ease of fixing CIs
  - fix both severe problems & make the easy fixes
- Will thinking aloud give the right answers?
  - not always
  - if you ask a question, people will always give an answer, even it is has nothing to do with facts
    - panty hose example

Measuring Bottom-Line Usability

- Situations in which numbers are useful
  - time requirements for task completion
  - successful task completion
  - compare two designs on speed or # of errors
- Ease of measurement
  - time is easy to record
  - error or successful completion is harder
- Do not combine with thinking-aloud.

Analyzing the Numbers



- Example: trying to get task time <=30 min.
  - test gives: 20, 15, 40, 90, 10, 5
  - mean (average) = 30
  - median (middle) = 17.5

Analyzing the Numbers (cont.)

- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that typical value is between 5 & 55

Analyzing the Numbers (cont.)

Web Usability Test Results						
Participant #		**Time (minutes)**				
1		20				
2		15				
3		40				
4		90				
5		10				
6		5				
	number of participants	**6**				
	mean	**30.0**				
	median	**17.5**				
	std dev	**31.8**				
	standard error of the mean	**= stddev / sqrt (#samples)**			**13.0**	
	typical values will be mean +/- 2*standard error			**--> 4 to 56!**		
	what is plausible? = confidence (alpha=5%, stddev, sample size)	**25.4**	**--> 95% confident between 5 & 56**			

Analyzing the Numbers (cont.)

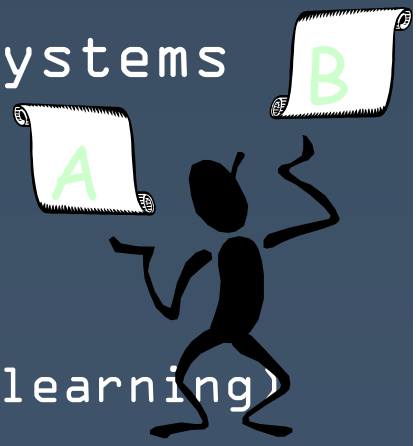
- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that typical value is between 5 & 55
- Usability test data is quite variable
  - need lots to get good estimates of typical values
  - 4 times as many tests will only narrow range by 2x
    - breadth of range depends on sqrt of # of test

Measuring User Preference

- How much users like or dislike the system
  - can ask them to rate on a scale of 1 to 10
  - or have them choose among statements
    - "best UI I've ever...", "better than average"...
  - hard to be sure what data will mean
    - novelty of UI, feelings, not realistic setting ...
- If many give you low ratings -> trouble
- Can get some useful data by asking
  - what they liked, disliked, where they

Comparing Two Alternatives

- *Between groups* experiment
  - two groups of test users
  - each group uses only 1 of the systems
- *Within groups* experiment
  - one group of test users
    - each person uses both systems
    - can't use the same tasks or order (learning)
  - best for low-level interaction techniques
- Between groups requires many more participants than within groups
- See if differences are statistically significant
  - assumes normal distribution & same std.



Experimental Details

- Order of tasks
  - choose one simple order (simple -> complex)
    - unless doing within groups experiment
- Training
  - depends on how real system will be used
- What if someone doesn't finish
  - assign very large time & large # of errors
- Pilot study
  - helps you fix problems with the study
  - do 2, first with colleagues, then with real users

Instructions to Participants

- Describe the purpose of the evaluation
  - "I'm testing the product; I'm not testing you"
- Tell them they can quit at any time
- Demonstrate the equipment
- Explain how to think aloud
- Explain that you will not provide help
- Describe the task
  - give written instructions, one task at

Details (cont.)

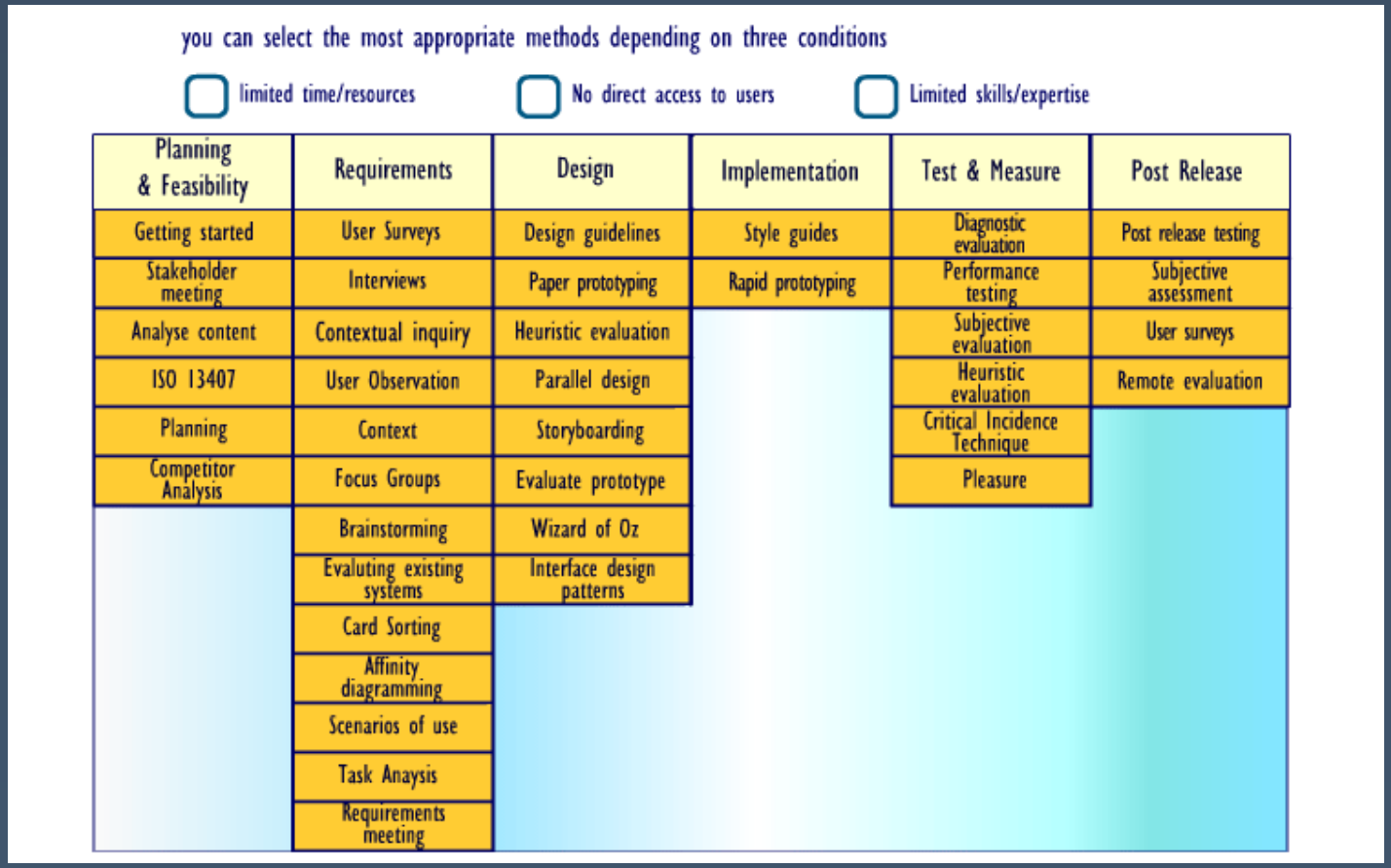
- Keeping variability down
  - recruit test users with similar background
  - brief users to bring them to common level
  - perform the test the same way every time
    - don't help some more than others (plan in advance)
  - make instructions clear
- Debriefing test users
  - often don't remember, so demonstrate or show video segments
  - ask for comments on specific features
    - show them screen (online or on paper)

Reporting the Results

- Report what you did & what happened
- Images & graphs help people
- Vi
co

Evaluation Methods

http://www.usabilitynet.org/tool

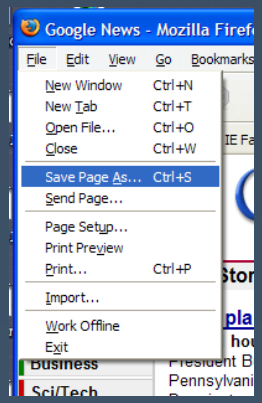


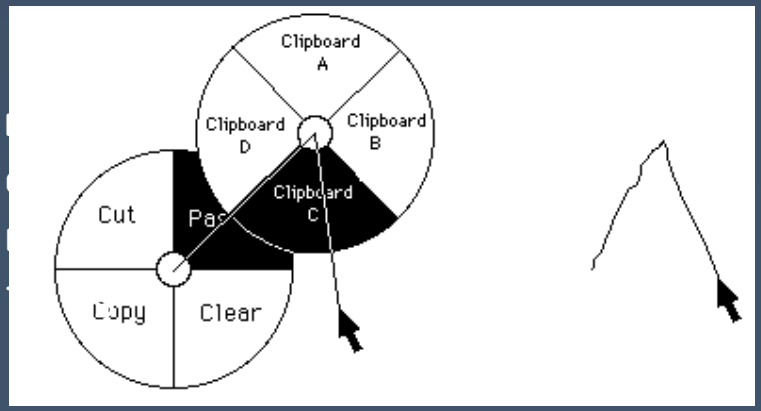
Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of lear
- Informality of Interacti
- Better Mappings/Embodime
- Agents - Proactive Compu
- Awareness
- Safety





http://www.billbuxton.com/PieMenus.html

Possible Benefits

- Efficiency
- **More User Control**
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

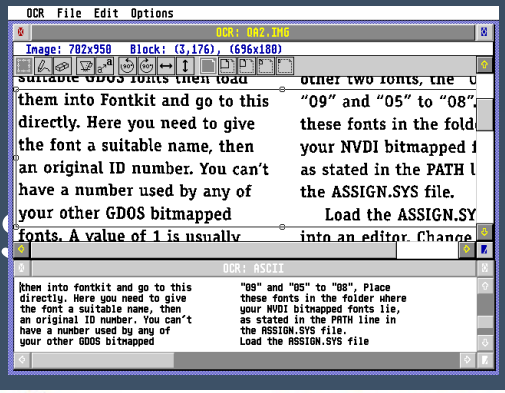
```
…
import ClientCookie openfun =
ClientCookie.urlopen reqfun =
ClientCookie.Request cj =
ClientCookie.LWPCookieJar()
if os.path.isfile(COOKIEFILE):
cj.load(COOKIEFILE)
opener=
ClientCookie.build_opener(ClientCookie.HTTPCoo
kieProcessor(cj))
ClientCookie.install_opener(opener)
```

Possible Benefits

- Efficiency
- More User Control
- **Affords Collaboration**
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learnir
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computir
- Awareness
- Safety

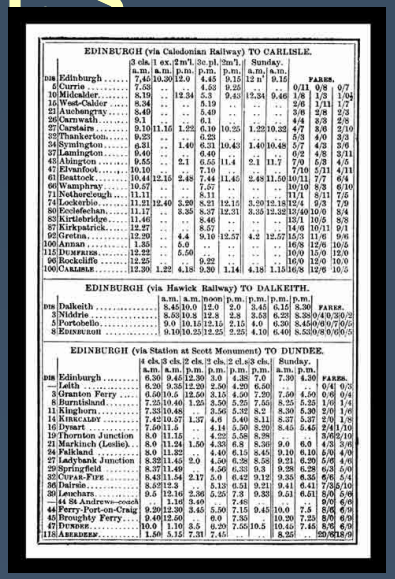
Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Acce
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety



Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- **See, Feel, Sense**
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety



EDINBURGH (via Caledonian Railway) TO CARLISLE.

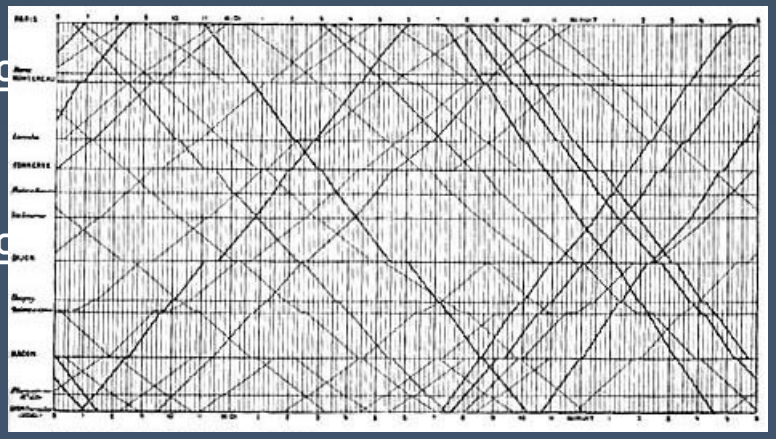
		3 cls.	1 ex.	2m'l.	3c. pl.	2m'l.	Sunday.		
		a.m.	a.m.	p.m.	p.m.	p.m.	a.m.	a.m.	
DIS	Edinburgh	7.45	10.30	12.0	4.45	9.15	12 n'	9.15	FARES.
5	Currie	7.59	..	..	4.59	9.29	..	..	0/11 0/8 0/7
10	Midcalder	8.19	..	12.34	5.3	9.43	12.34	9.46	1/8 1/3 1/04
16	West-Calder	8.34	..	..	5.19	..	..	..	2/5 1/11 1/7
21	Auchengray	8.49	..	..	5.49	..	..	..	3/6 2/8 2/3
26	Carnwath	9.1	..	..	6.1	..	..	..	4/4 3/3 2/8
27	Carstairs	9.10	11.10	1.22	6.10	10.25	1.22	10.32	4/7 3/6 2/10
29	Thankerton	9.23	..	..	6.23	..	..	..	5/3 4/0 3/3
34	Symington	9.31	..	1.40	6.31	10.43	1.40	10.48	5/7 4/3 3/6
37	Lamington	9.40	..	..	6.40	..	..	..	6/2 4/8 3/11
43	Abington	9.55	..	2.1	6.55	11.4	2.1	11.7	7/0 5/3 4/5
47	Elvanfoot	10.10	..	..	7.10	..	..	..	7/10 5/11 4/11
61	Beattock	10.44	12.15	2.48	7.44	11.45	2.48	11.50	10/11 7/7 6/4
66	Wamphray	10.57	..	..	7.57	..	..	..	10/10 8/3 6/10
71	Nethercleugh	11.11	..	..	8.11	..	..	..	11/4 8/11 7/5
74	Lockerbie	11.21	12.40	3.39	8.21	12.16	3.39	12.18	12/4 9/3 7/9
80	Ecclefechan	11.17	..	3.35	8.57	12.31	3.35	12.32	12/40 10/0 8/4
83	Kirtlebridge	11.46	..	..	8.46	..	..	..	13/1 10/5 8/8
87	Kirkpatrick	12.27	..	..	8.57	..	..	..	14/0 10/11 9/1
92	Gretna	12.39	..	4.4	9.10	12.57	4.2	12.57	15/3 11/6 9/6
100	Annan	1.33	..	5.0	..	..	..	..	16/8 12/6 10/5
115	Dumfries	12.22	..	5.50	..	..	..	..	16/0 15/0 12/0
96	Rockcliffe	12.35	..	..	9.22	..	..	..	16/0 12/0 10/0
100	Carlisle	12.39	1.22	4.18	9.30	1.14	4.18	1.15	16/8 12/6 10/5

EDINBURGH (via Hawick Railway) TO DALKEITH.

		a.m.	a.m.	noon	p.m.	p.m.	p.m.	p.m.	
DIS	Dalkeith	8.45	10.0	12.0	2.0	3.45	6.15	8.30	FARES.
5	Niddrie	8.55	10.3	12.3	2.3	3.53	6.23	8.39	0/4 0/3 0/2
5	Portobello	9.0	10.15	12.15	2.15	4.0	6.30	8.45	0/6 0/5 0/3
8	Edinburgh	9.10	10.25	12.25	2.25	4.10	6.40	8.55	0/8 0/6 0/3

EDINBURGH (via Station at Scott Monument) TO DUNDEE.

		4 cls.	3 cls.	2 cls.	2 cl. s.	3 cls.		Sunday.		
		a.m.	a.m.	p.m.	p.m.	p.m.	p.m.	a.m.	p.m.	
DIS	Edinburgh	6.30	9.45	12.30	3.0	4.38	7.0	7.30	4.30	FARES.
—	Leith	6.39	9.55	12.39	3.50	4.39	6.50	..	..	0/4 0/3
3	Granton Ferry	6.50	10.5	12.50	3.15	4.50	7.20	7.50	4.50	0/6 0/4
8	Burntisland	7.25	10.40	1.25	3.50	5.25	7.50	8.25	5.25	1/0 1/4
11	Kinghorn	7.33	10.48	..	3.56	5.32	8.2	8.30	5.30	2/0 1/6
14	Kirkcaldy	7.42	10.57	1.37	4.6	5.40	8.11	8.37	5.37	2/0 1/8
16	Dysart	7.50	11.5	..	4.14	5.50	8.20	8.45	5.45	2/4 1/10
19	Thornton Junction	8.0	11.15	..	4.23	5.58	8.38	..	..	3/0 2/0
21	Markinch (Leslie)	8.0	11.24	1.50	4.33	6.8	8.36	9.0	6.0	4/3 3/0
24	Falkland	8.0	11.32	..	4.40	6.15	8.45	9.10	6.10	5/0 4/0
27	Ladybank Junction	8.32	11.45	2.0	4.50	6.38	8.58	9.21	6.20	5/6 4/6
29	Springfield	8.37	11.49	..	4.56	6.33	9.3	9.28	6.28	6/3 5/0
32	Cupar-Fife	8.43	11.54	2.17	5.0	6.42	9.12	9.35	6.35	6/6 5/4
36	Dairsie	8.53	12.3	..	5.13	6.51	9.21	9.41	6.41	7/3 5/10
38	Leuchars	9.6	12.16	2.36	5.26	7.3	9.33	9.51	6.51	8/0 6/6
—	44 St Andrews—coach	..	1.16	3.40	..	7.46	..	..	..	9/6 6/6
44	Ferry-Port-on-Craig	9.20	12.30	3.45	5.50	7.15	9.45	10.0	7.5	8/6 6/6
45	Broughty Ferry	9.40	12.50	..	6.9	7.35	..	10.20	7.25	8/6 6/6
47	Dundee	10.0	1.10	4.5	6.20	7.55	10.5	10.45	7.45	8/6 6/6
118	Aberdeen	..	1.50	5.15	7.31	7.45	..	..	8.25	20/6 18/6



E J Marey, Paris to Lyon Train Schedule, 1878

Possible Benefits

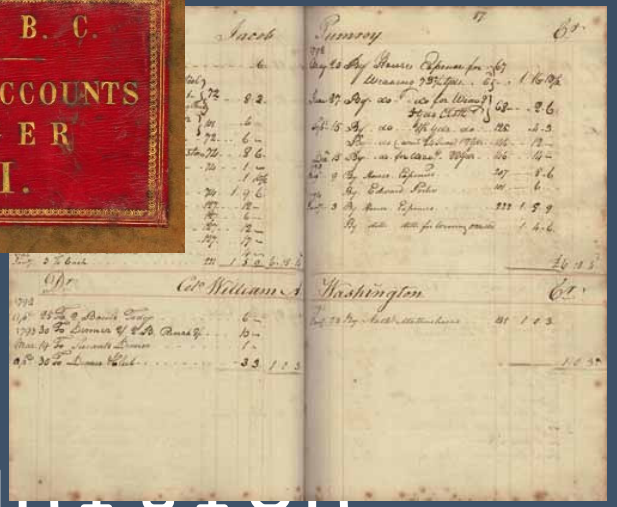
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Org
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

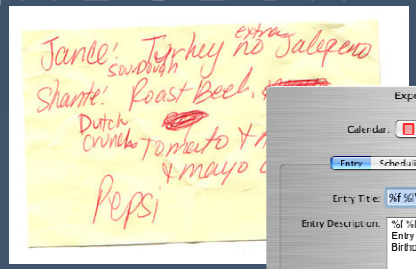
Possible Benefits

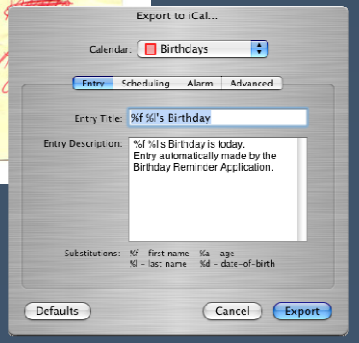
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- **Communication**
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cog...
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

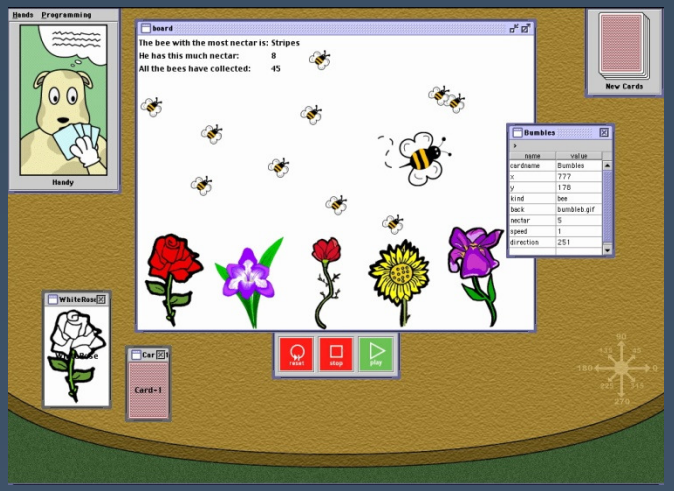






Possible Benefits

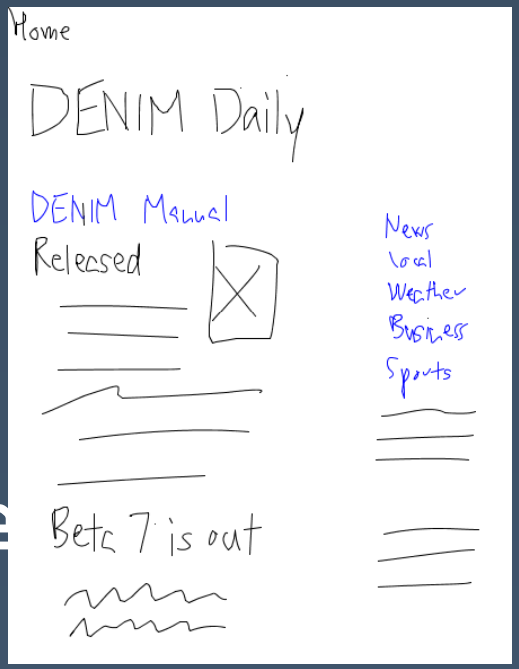
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition



- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Inte
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety



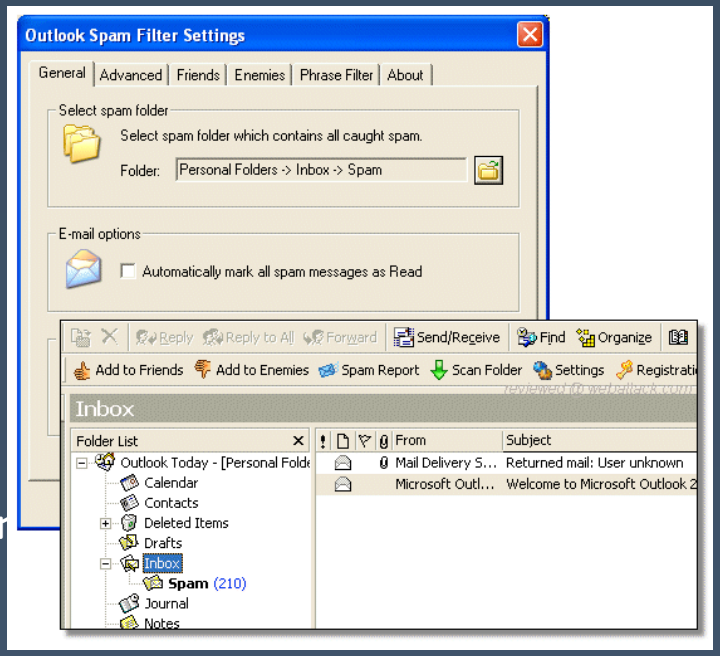
Mark Newman et al,
DENIM

Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- **Better Mappings/Embodiment**
- Agents - Proactive Computing
- Awareness
- Safety

Possible Benefits

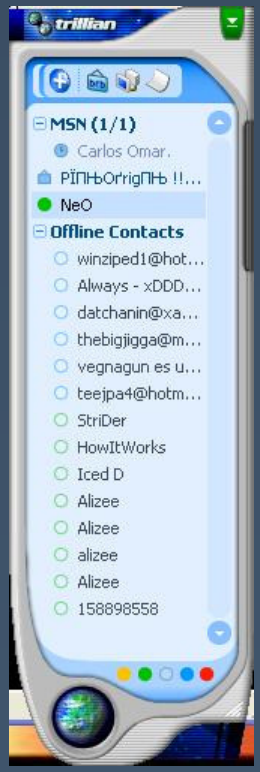
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learnir
- Informality of Interaction
- Better Mappings/Embodiment



- Agents - Proactive Computing
- Awareness
- Safety

Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- **Awareness**
- Safety



Possible Benefits

- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents - Proactive Computing
- Awareness
- Safety

Some of these slides are borrowed from James Landay, Jason Hong, & Scott Klemmer's Berkeley summer HCI course materials

In-class exercise (Kevin and Nundu)

- Find another group to pair up with (ideally, one with a different mentor)
- Group A gives elevator pitch
- Group B synthesizes A's design problem and writes it down (in idea log)
- then switch

5 minutes

Split into sections… and, into pairs of groups