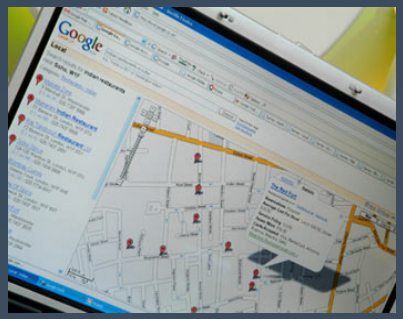
Evaluation

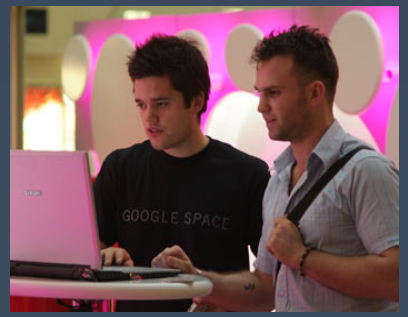
Scott Klemmer

TAs: Marcello Bastea-Forte, Joel Brandt, Neil Patel, Leslie Wu, Mike Cammarano

13 November 2007

How can we measure success?





Google “airport lounge” / user testing lab in Heathrow airport

Source: ZDNet

Poor tests yield poor results

Friday, August 17, 2007 11:03 AM PT Posted by Harry McCracken

A Not-Very-Useful iPhone Keyboard Study

ADD TO MY PAGES PRINT E-MAIL COMMENT RSS

SLASHDOT IT DIGG THIS DEL.ICIO.US NEWSVINE

Research firm User Centric has released a study that tries to gauge how effective the iPhone's unusual on-screen keyboard is. The goal is certainly a noble one, but I can't say that the survey's approach results in data that makes much sense.

User Centric brought in twenty owners of other phones--half who had ones with QWERTY keyboards, and half who had ordinary numeric phone keypads. None were familiar with the iPhone. The research involved having the test subjects enter six sample text messages with the phones they already had, and six with an iPhone.

Logical end result: These iPhone newbies took twice as long to enter text with an iPhone as they did with their own phones, and made lots more typos.

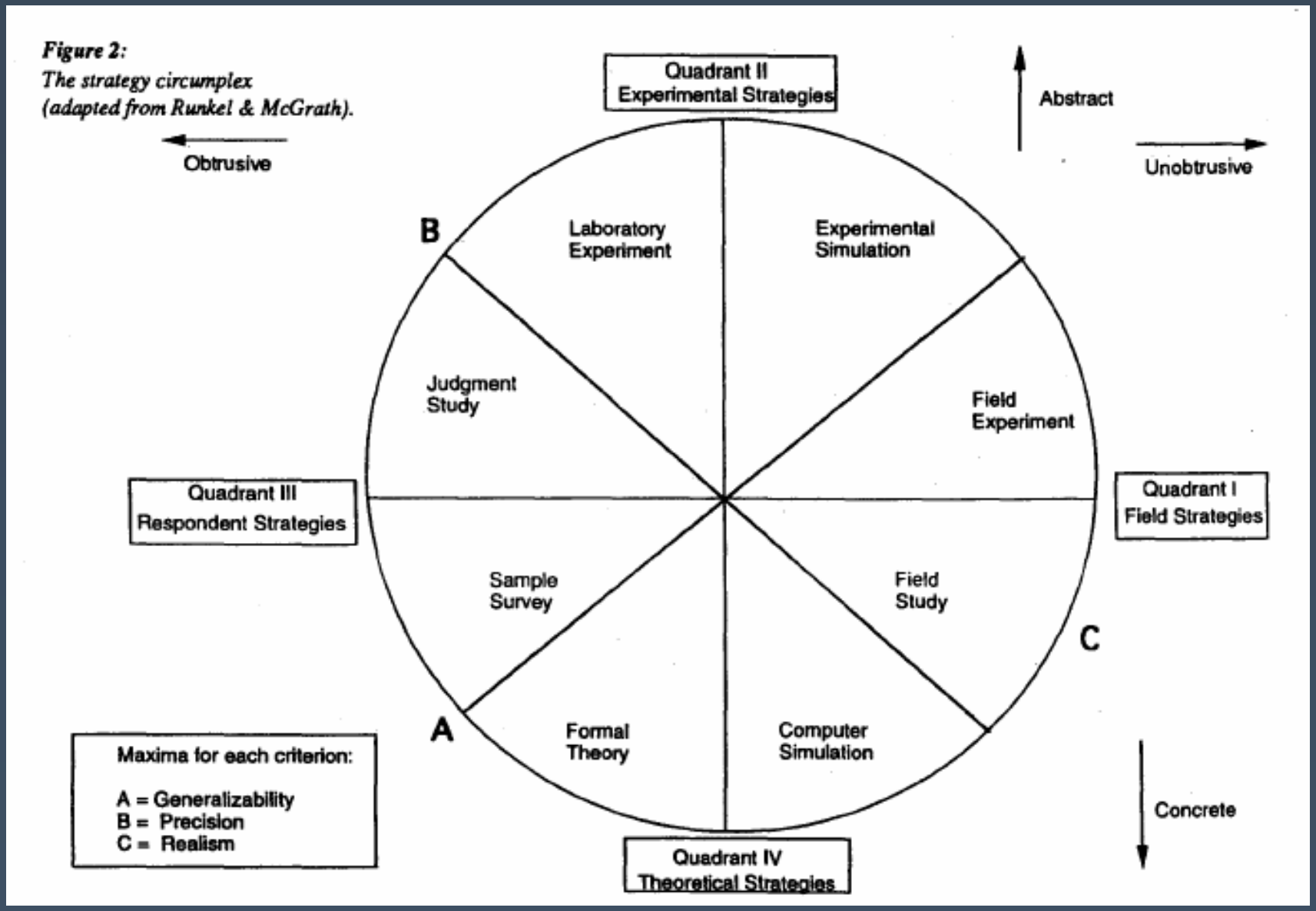
Issues

- user sample
- statistical significance
- "newbie" effect / learning effects

One interesting result: iPhone did same or better for multitap users

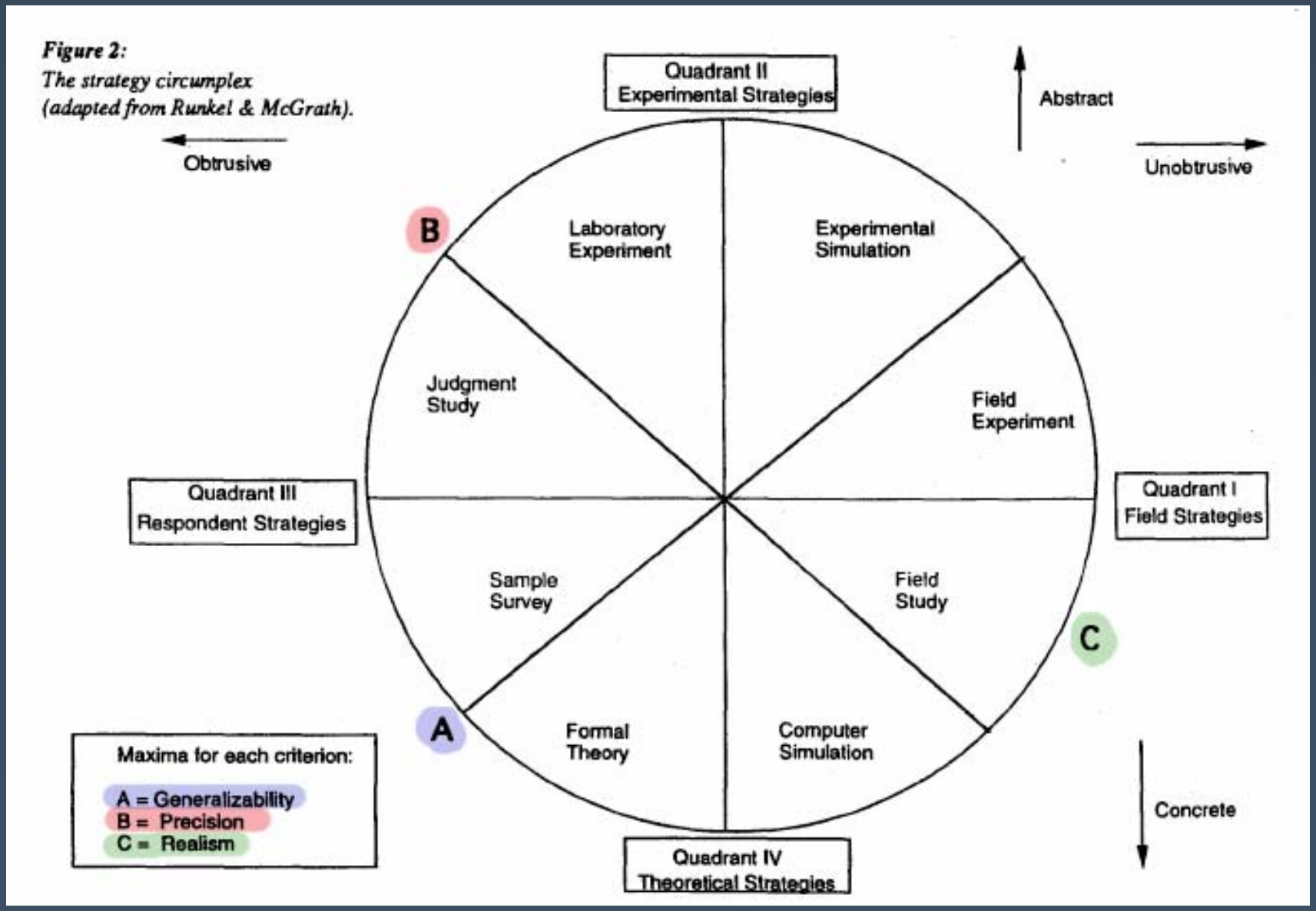
Source: PC World

Taxonomy of Methods



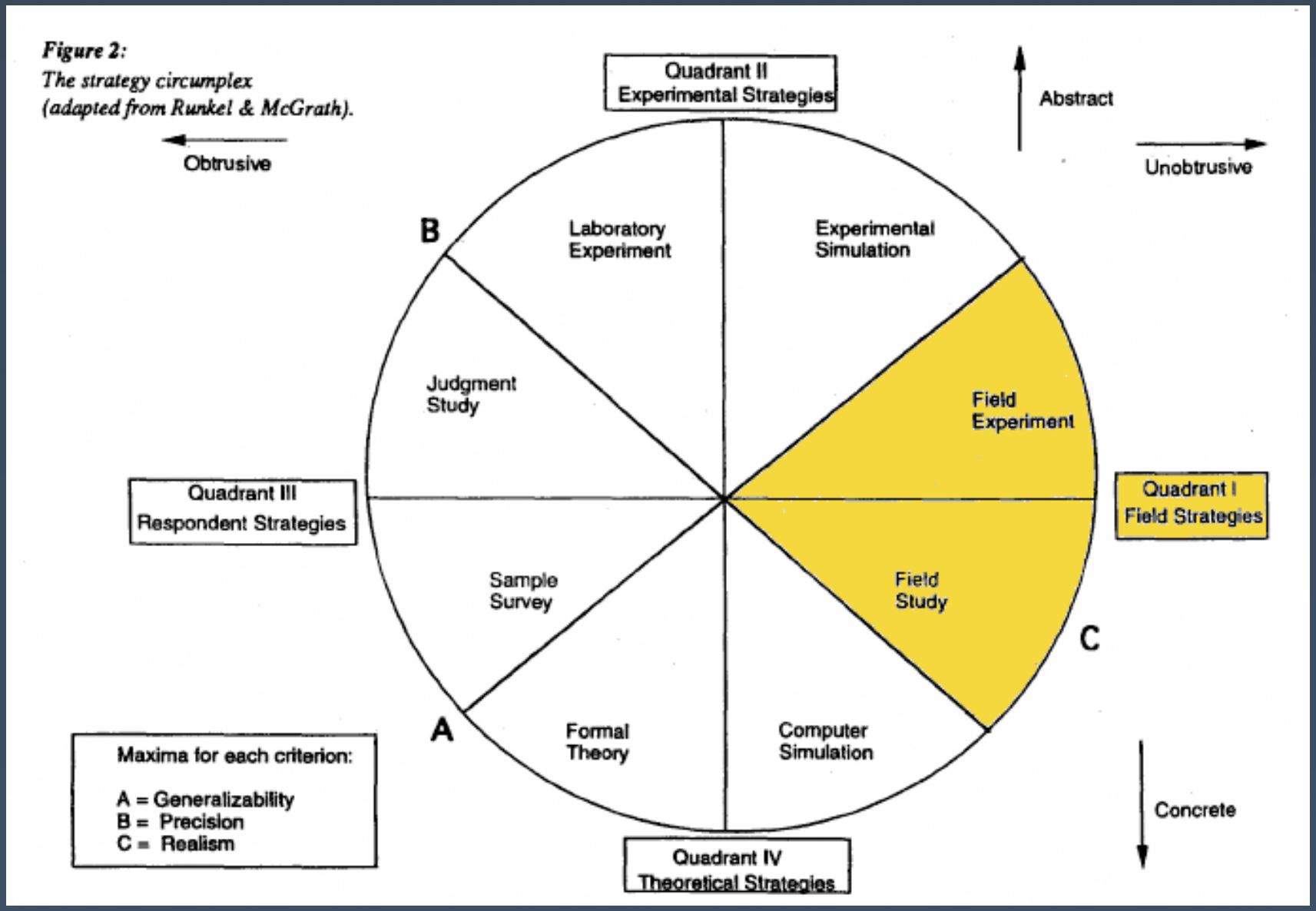
McGrath et al. 1994

Taxonomy of Methods



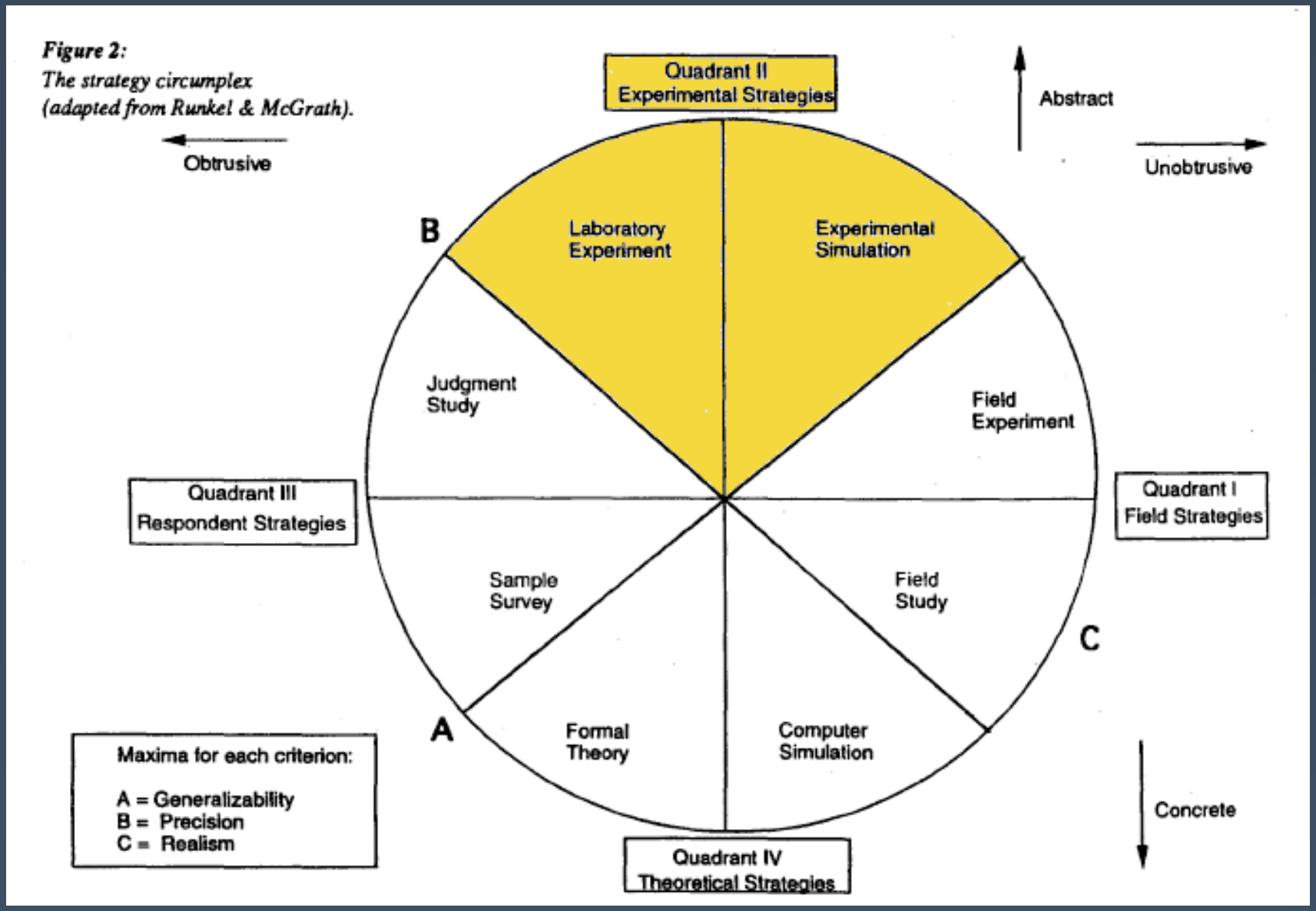
McGrath et al. 1994

Taxonomy of Methods



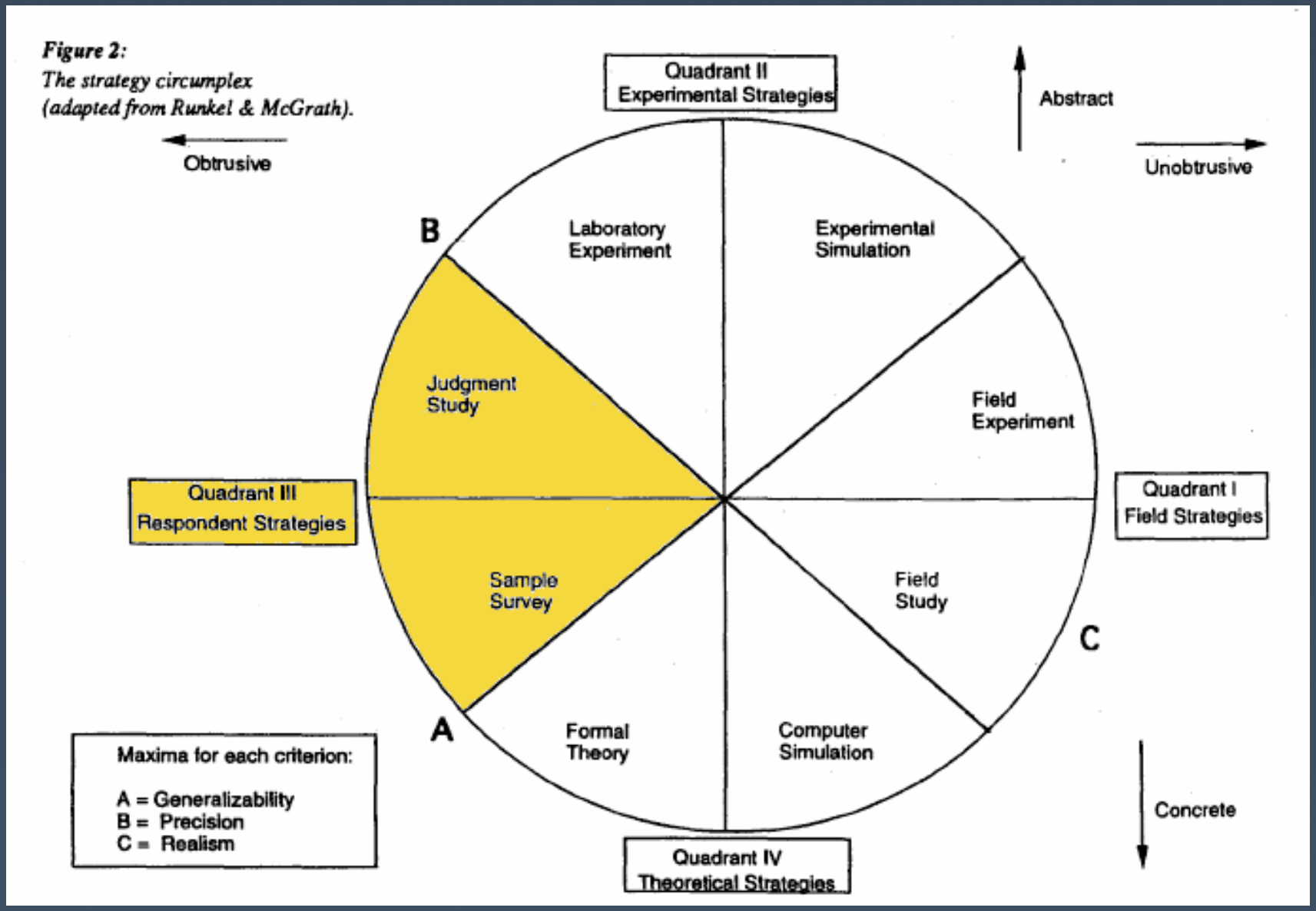
McGrath et al. 1994

Taxonomy of Methods



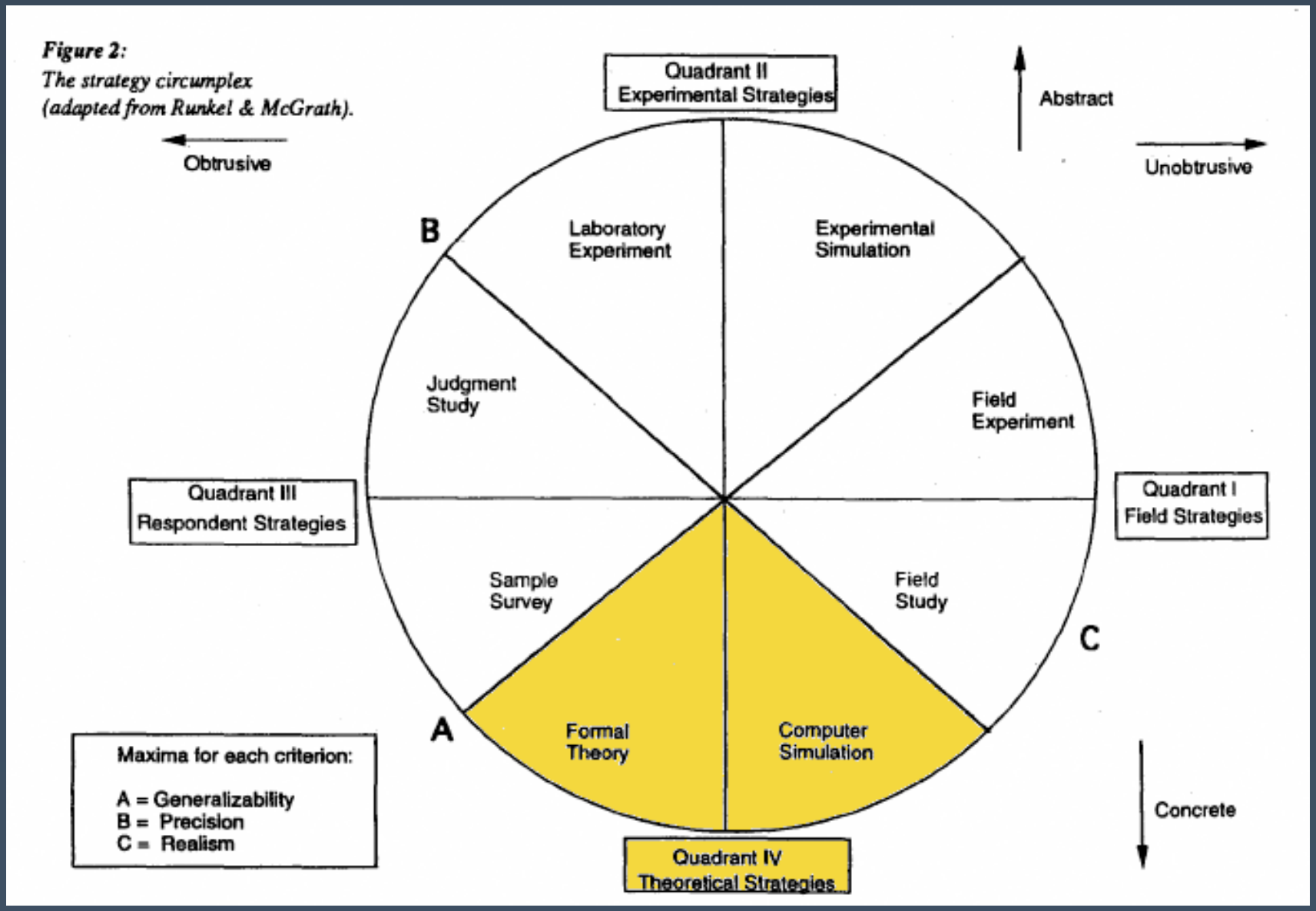
McGrath et al. 1994

Taxonomy of Methods



McGrath et al. 1994

Taxonomy of Methods



McGrath et al. 1994

What is usability testing?

- **Usability testing** is a means for measuring how well people can use some human-made object (such as a web page, a computer interface, a document, or a device) for its intended purpose, i.e. usability testing measures the usability of the object.

Source: Wikipedia

Why do User Testing?

- Can't tell how good UI is until?
  -
- 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

Why do user studies?

Motivation

- identify any usability problems that the product has
- collect quantitative data on participants' performance
- determine participants' satisfaction with the product

Source: Usability.gov

Empirical Questions

- Baserates: How often does Y occur?
  - Requires measuring Y.
- Correlations: Do X and Y co-vary?
  - Requires measuring X and Y.
- Causes: Does X cause Y?
  - Requires measuring X and Y, and manipulating X.
  - Also requires somehow accounting for the effects of other independent variables (confounds)!

What will you measure?

- **Time on Task** -- How long does it take people to complete basic tasks? (For example, find something to buy, create a new account, and order the item.)
- **Accuracy** -- How many mistakes did people make? (And were they fatal or recoverable with the right information?)
- **Recall** -- How much does the person remember afterwards or after periods of non-use?
- **Emotional Response** -- How does the person feel about the tasks completed? (Confident? Stressed? Would the user recommend this system to a friend?)

Source: Wikipedia

Terminology

- Response variables (a.k.a. *dependent* variable(s))
  - Outcomes of experiment
- Factors (a.k.a. *independent* variables))
  - Variables we manipulate in each condition
- Levels (a.k.a. possible values for independent variables)
- Replication (number of subjects assigned to each level)

Goals

- Internal validity
  - Manipulation of independent variable is cause of change in dependent variable
    - Requires removing effects of confounding factors
    - Requires choosing a large enough sample size, so the result couldn't have happened by chance alone.
- External validity
  - Results generalize to real world situations
  - Requires that the experiment be replicable
  - No study "has" external validity by itself!

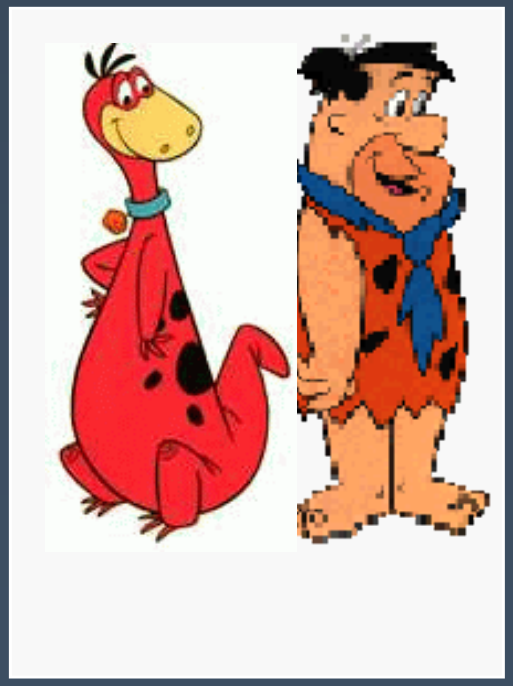
Control vs. Randomization

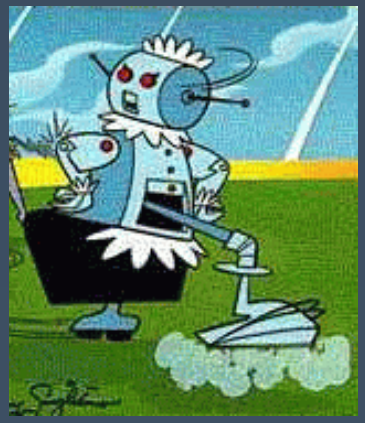
- Control: holding a variable constant for all cases
  - Lower generalizability of results
  - Higher precision of results
- Randomization: allowing a variable to randomly vary for all cases
  - Higher generalizability of results
  - Lower precision of results
- Randomization within blocks: allowing a variable to randomly vary with some constraints
  - Compromise approach

Between subjects design

- Wilma and Betty use one interface

Dino and Fred use the other



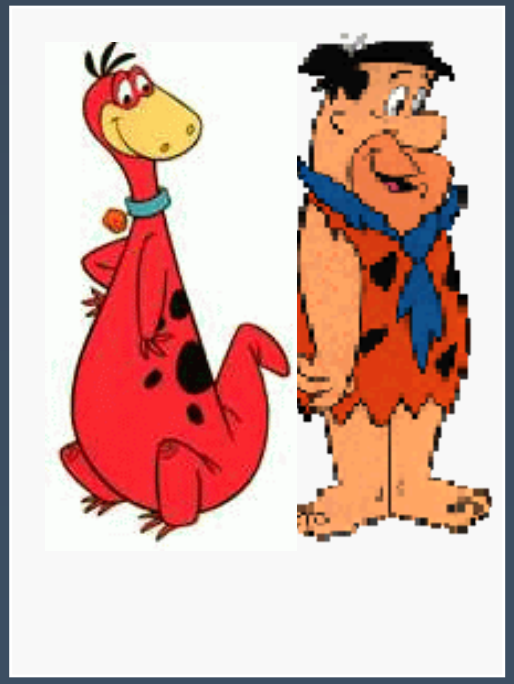


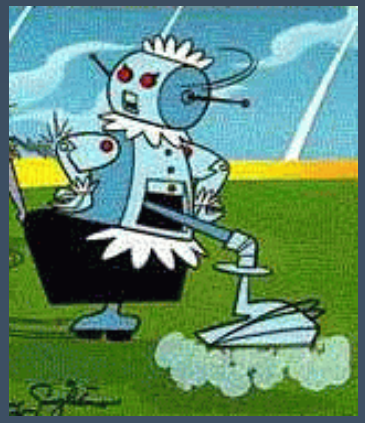
Between subjects design

Subjects wearing 7-*mm* spikes	Time (in seconds)	Subjects wearing 13-*mm* spikes	Time (in seconds)
Mike	11.7	Don	15.7
Bob	18.2	Hector	13.4
Homer	12.2	Ron	18.0
George	15.4	Tom	12.8
Harry	15.8	Steve	13.6
Gordon	13.2	Dale	19.0
John	13.7	Pete	16.2
Bill	19.1	Juan	11.9
Randy	12.9	Dan	14.6
Tim	16.0	Paul	18.0

Within subjects design

- Everyone uses both interfaces





Within subjects design

Subjects on manual typewriter	Typing speed (wpm)	Subjects on electric typewriter	Typing speed (wpm)
Mike	35	Mike	40
Bob	42	Bob	45
Homer	32	Homer	35
George	25	George	30
Harry	30	Harry	32
Gordon	30	Gordon	35
John	30	John	40
Bill	36	Bill	37
Randy	36	Randy	42
Tim	30	Tim	34

Ordering effects

- Ordering of conditions is a variable that can confound the results

  - Randomization
  - Counterbalancing
  - Latin square (partial counterbalancing)

- "The day should come then when no reputable psychologist will use a within-subject design... without combining it with a separate groups [between subjects] design." - Poulton 1973

Between vs. within subjects

- Within subjects
  - All participants try all conditions
    - + Can isolate effect of individual differences
    - + Requires fewer participants
    - - Ordering and fatigue effects
- Between subjects
  - Each participant tries one condition
    - + No ordering effects, less fatigue.
    - - Cannot isolate effects due to individual differences.
    - - Need more participants

Choosing Participants

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

What should you keep in mind?

- You are testing the site not the users.
- Rely more on what you learn about performance than preference.
- Make use of what you learn.
- Try to find the best solution given the reality of your many users.

- Follow University Human Subject Guidelines

Source: Usability.gov

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

Setting goals / developing the test plan

Scope

- What are you testing?

Purpose

- What concerns, questions, and goals is the test focusing on?

Schedule and location

- When and where will the test take place?

Participants

- How many users of what types will you recruit?

Scenarios

- What will participants do with the product in this round of testing?

Questions

- What will you ask at the beginning and end of the session?

Data to be collected

- What will you count?

Set up

- What system will you use for testing? Will you be videotaping and/or audiotaping? Will you be using a specific technology to capture data?

Roles

- Who will do what in the usability test?

Source: Usability.gov

Creating final plan

- A good plan for usability testing gives the participants:
  - a goal/task (what to do or what question to find the answer for)
  - data, if needed, that a real user would have when going to the site to do that task
- You can give the scenario as just the statement of the goal/task or you can elaborate it a little with a very short story that adds motivation to get to the goal.

Source: Usability.gov

Recruiting participants

- The participants must be like the people who will use your site.
- You might end up using a screening questionnaire
- … plan on a cost associated with finding the people … you may still need to plan on incentives to get participants to participate …

Source: Usability.gov

Setting up the test sessions

- Make sure you have everything you need
  - the prototype you are going to test
  - the computer set up for the participant with the monitor, resolution, and connection speed that you indicated in the test plan
  - note-taking forms on paper or set up on a computer
  - consent forms for participants to sign and a pen in case the participant does not bring one
  - questionnaires, if you are using any
  - the participant's copy of the scenarios
  - cameras, microphones, or other recording equipment if you are using any
  - folders to keep each person's paperwork in if you are using paper
- Do a dry-run and a pilot test

Source: Usability.gov

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
- Don't choose tasks that are too fragmented
  - e.g., phone-in bank test

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

Summary: Before Starting

- You should know, and have written down
  - objective
  - description of system being testing
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures

- Seems tedious, but writing this will help “debug” your test

Conducting the usability test

- The facilitator:
  - welcomes the participant and introduces anyone else who is in the room
  - invites the participant to sit in front of the computer where the participant will be working
  - explains the general goal of the session—to have the participant try out a Web site (or whatever the product is that is being tested)
  - asks participant profile questions and has the participant sign the release form
  - explains thinking aloud (and may demonstrate it and have the participant do a think aloud exercise)
  - asks if the participant has any questions before starting and answers any that will not give away what you want to learn from the participant
  - tells the participant where to start
- The participant starts to work with the Web site (or other product).
  - The participant works on the scenario while thinking aloud. The note-takers take notes.
  - The session continues from scenario to scenario until the participant has done (or tried) them all or the time allotted has elapsed.
- The facilitator asks the end-of-session questions
  - thanks the participant, giving the participant the agreed-on incentive, and escorts the participant out.

Source: Usability.gov

Options for capturing results

- Think aloud
- Nothing critical incidents
- Video recording
- Screen recording

- Decide whether to interrupt or not
- Keep variances among tests low

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)

Analyzing the results

- Quantitative data, which might include:
  - success rates
  - time to complete tasks
  - pages visited
  - error rates
  - ratings on a satisfaction questionnaire
- Qualitative data, which might include:
  - notes of your observations about the pathways participants took
  - notes about problems participants had (critical incidents)
  - notes of what participants said as they worked
  - participants' answers to open-ended questions

Source: Usability.gov

Using the Test Results

- Summarize the data
  - make a list of all critical incidents
    - 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?
  - something missing?

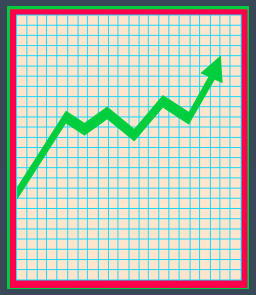
Using the Results (cont.)

- 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
  - try to avoid specific questions

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
    - define in advance what these mean
- Do not combine with thinking-aloud. Why?
  - t

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
  - looks good!

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 users
  - this is when online methods become useful
    - easy to test w/ large numbers of users

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 had trouble, best part, worst part, etc. (redundant questions are OK)

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 a time

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 get it!
- Video clips can be quite convincing

Bruce Tognazzini

- **advocates close-coupled testing**: "Run a test subject through the product, figure out what's wrong, change it, and repeat until everything works."

Source: Wikipedia

More commentary from Tog

AskTog: Note created September 12, 2007

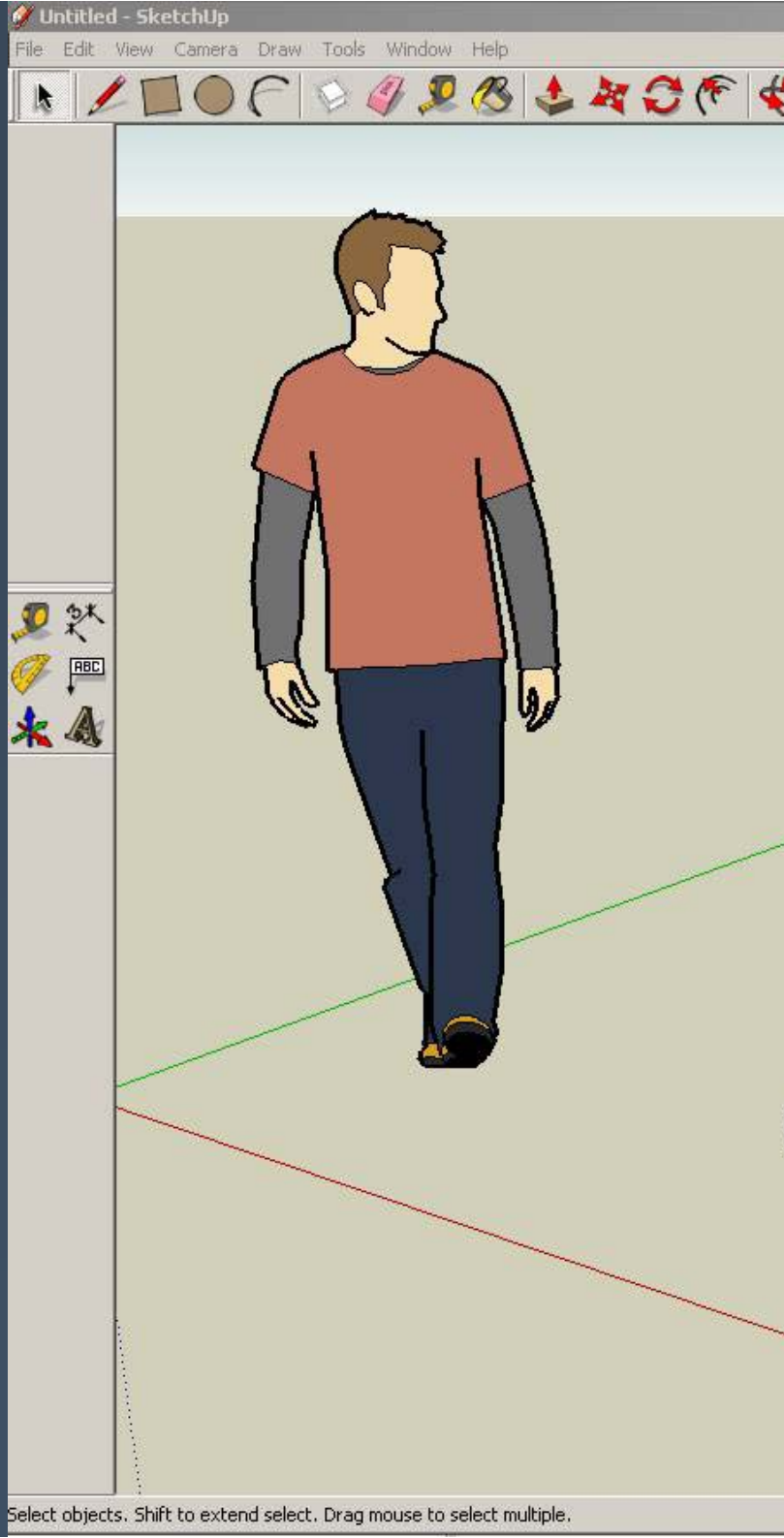
- I have spent much of my twenty-five year career in software design troubleshooting projects that are in trouble, helping them get back on course. The single thread that ran through every one of them was a lack of user testing.

- If your people don't know how to user test, find someone who does. Whether you bring in an outside design firm or hire your own people, make sure they are telling you all about their plans to test, because if they don't test, your customers will, and it will cost you a whole bunch more money.

Source: http://www.asktog.com/columns/037TestOrElse.html

CASE STUDY

David Akers evaluation of Google SketchUp



Study Goals

1. What individual differences (previous software used, computer use, spatial reasoning ability, etc.) best predict performance on simple modeling tasks?
2. What usage log metrics (e.g. frequency of undo operations, frequency of camera operations, etc.) best predict performance on simple modeling tasks?
3. What specific problem do novice SketchUp users encounter most frequently on simple modeling tasks?

n = 54

90% students

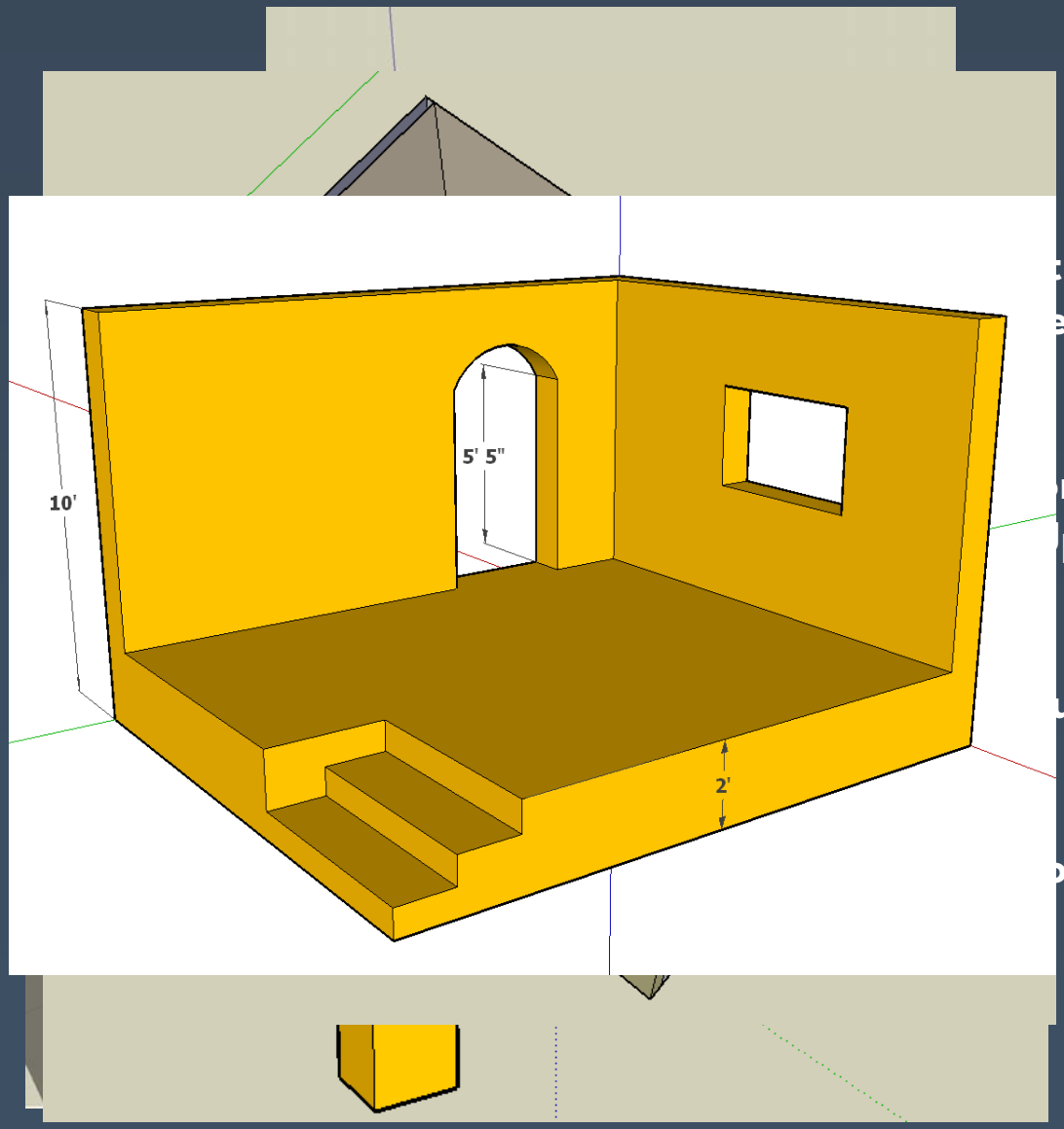
35% architecture
20% computer science
10% mechanical engineering
10% civil engineering
25% other (art, physics, etc.)

41% never used
44% novice
15% intermediate


SketchUp Usability Study – Boulder, CO
September, 2007
Participant Instruction Packet
SketchUp

Study Design

1. Entry questionnaire (5 min.)
2. Mental rotation test (15 min.)
3. Video tutorials (15 min.)
4. Free exploration (10 min.)
5. Tasks (3 x 15 min.)
6. Exit questionnaire (5 min.)



Data Size

Event log data (450 MB)

Screen capture video (75 GB!)

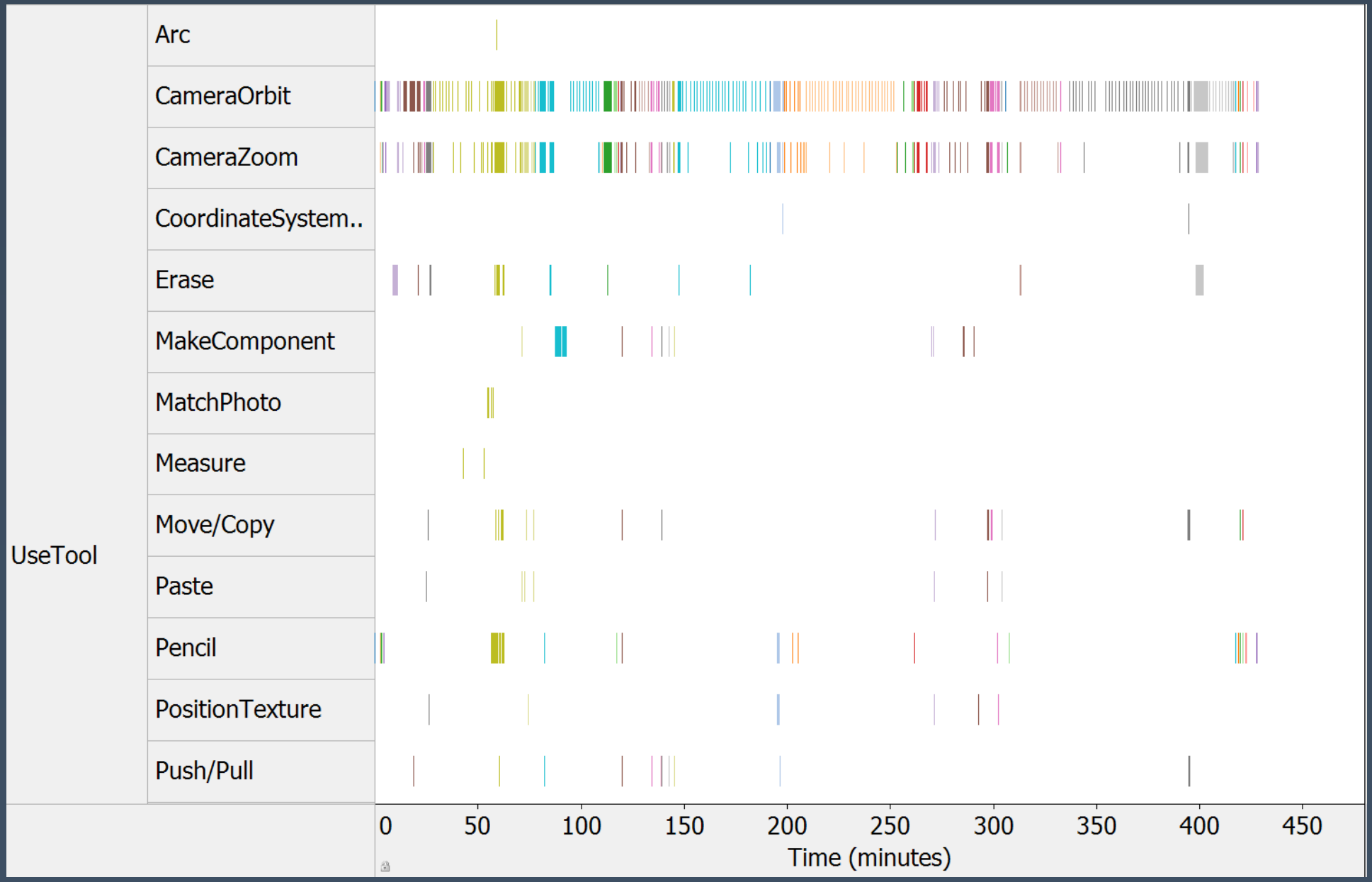
3D models (100 MB)

Questionnaires (17 KB)

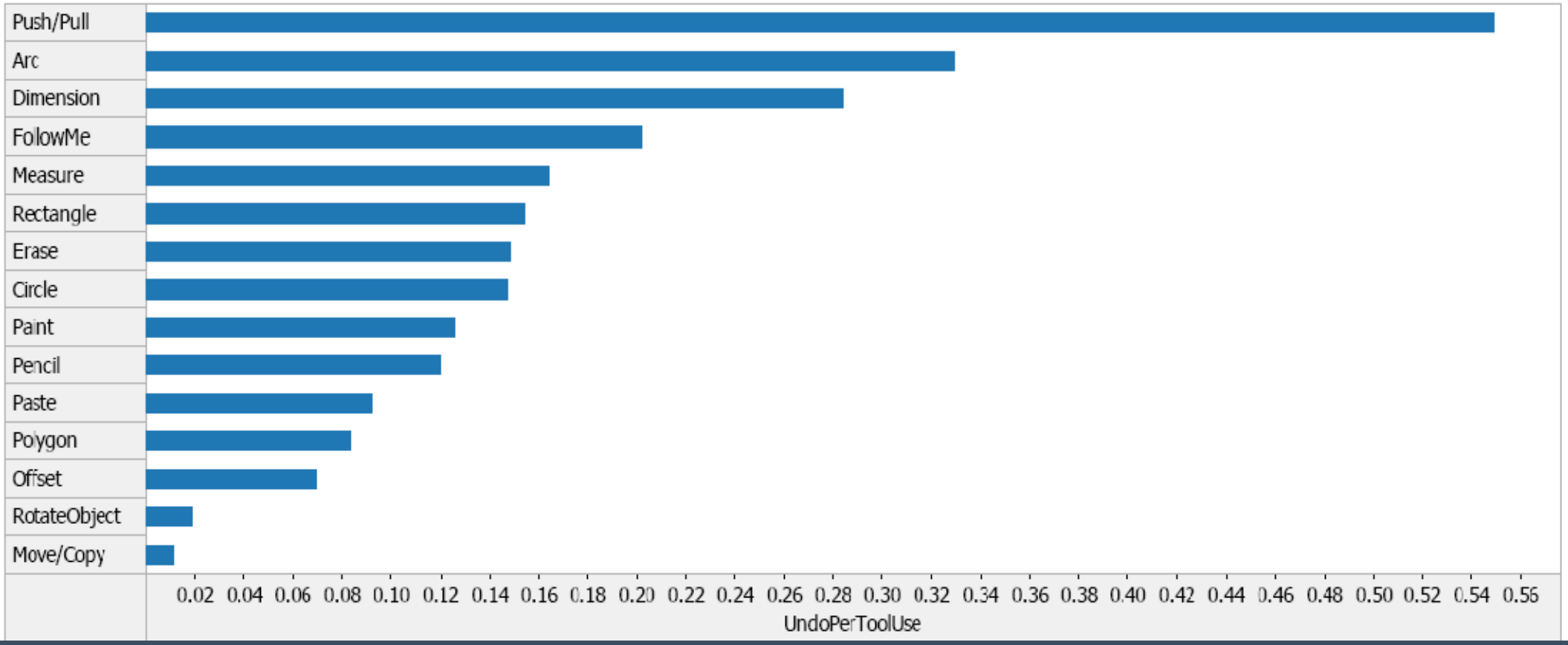
Study Goals

1. What individual differences (previous software used, computer use, spatial reasoning ability, etc.) best predict performance on simple modeling tasks?
2. What usage log metrics (e.g. frequency of undo operations, frequency of camera operations, etc.) best predict performance on simple modeling tasks?
3. **What specific problem do novice SketchUp users encounter most frequently on simple modeling tasks?**

Log Analysis of Tool Usage



Undo Rates


Push/Pull
Arc
Dimension
FolowMe
Measure
Rectangle
Erase
Circle
Paint
Pencil
Paste
Polygon
Offset
RotateObject
Move/Copy
0.02 0.04 0.06 0.08 0.10 0.12 0.14 0.16 0.18 0.20 0.22 0.24 0.26 0.28 0.30 0.32 0.34 0.36 0.38 0.40 0.42 0.44 0.46 0.48 0.50 0.52 0.54 0.56
UndoPerToolUse

Developer Hypotheses (Wrong)

For the Push/Pull tool:

90% of undo operations are caused by bugs in SketchUp.

10% of undo operations are caused by difficulties with inferencing.

EYE TO THE FUTURE

Instrumenting Applications

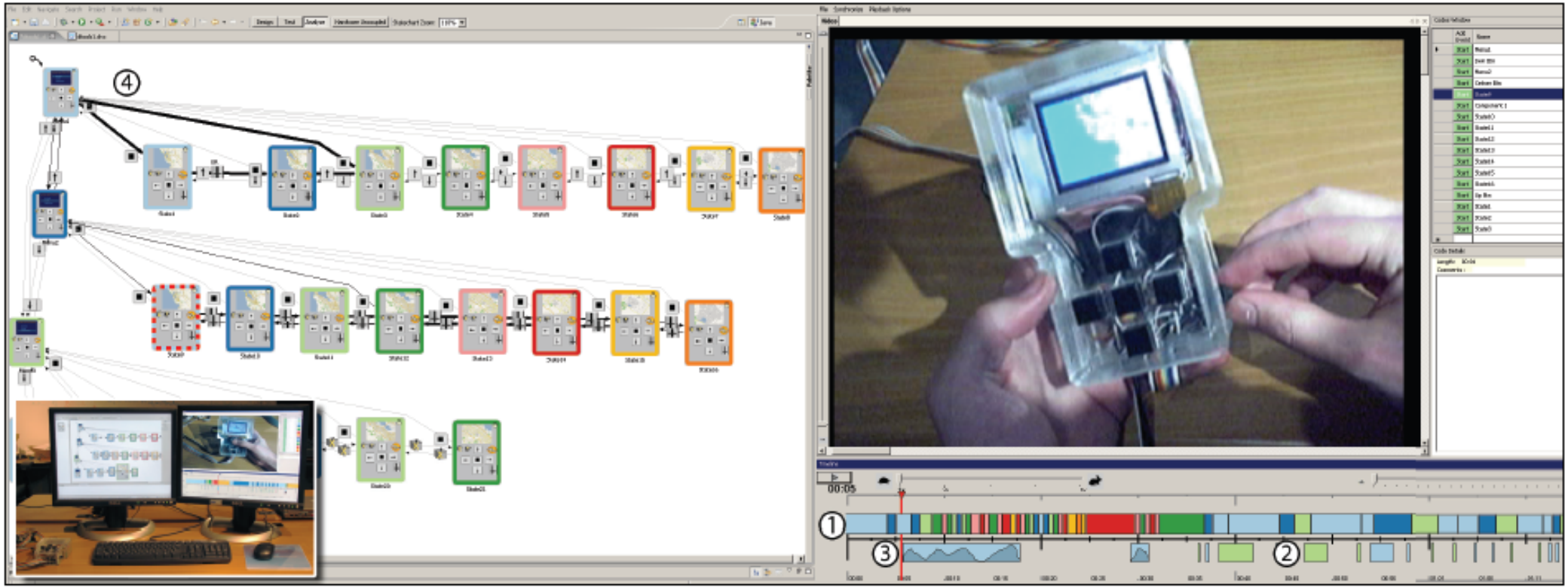


Figure 3 In Analysis mode, statechart and recorded video are synchronized and each can be used to access the other. *Inset:* simultaneous interaction with statechart and video editing is possible on a dual-screen workstation.

d.Tools physical prototyping captures user tests

Source: Hartmann, B., Klemmer, S.R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., Gee, J. Reflective physical prototyping through integrated design, test, and anal
Proceedings of UIST 2006, October 2006

Thanks to...

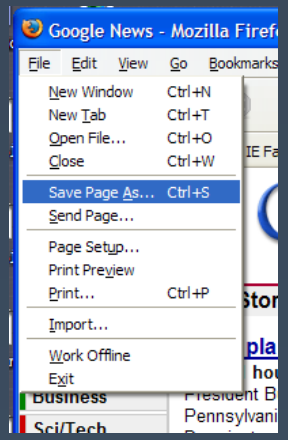
- David Akers, Anoop Sinha, and James Landay for contributing materials to these slides

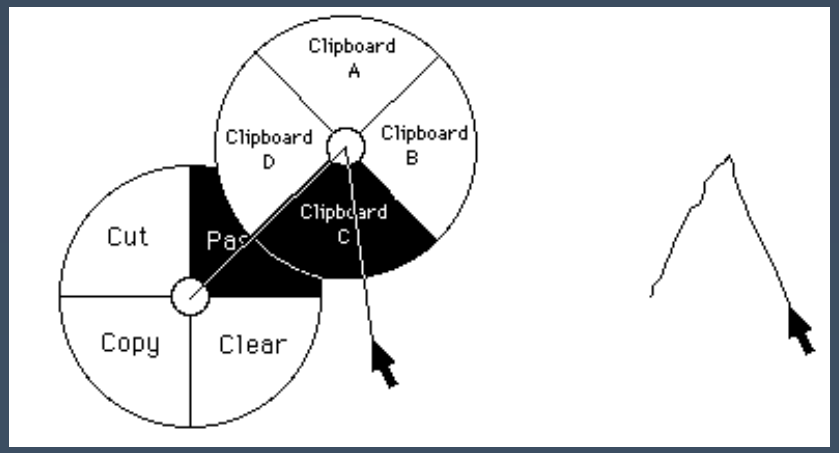
Possible Benefits

- 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





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)
```

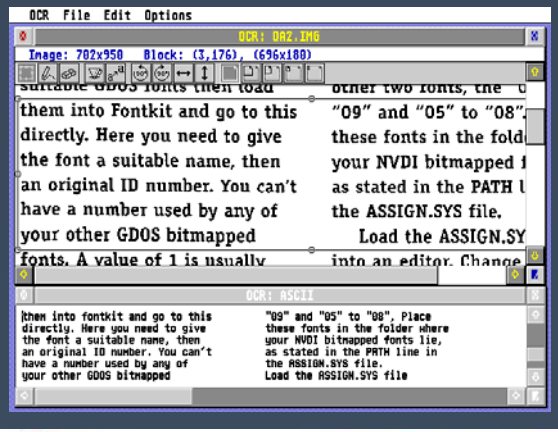
…

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

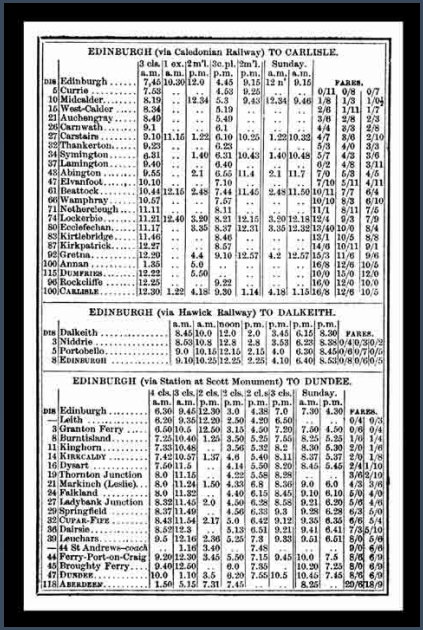
Possible Benefits

- 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 Interaction
- Better Mappings/Embodiment
- Agents – Proactive Computing
- Awareness
- Safety



EDINBURGH (via Caledonian Railway) TO CARLISLE.

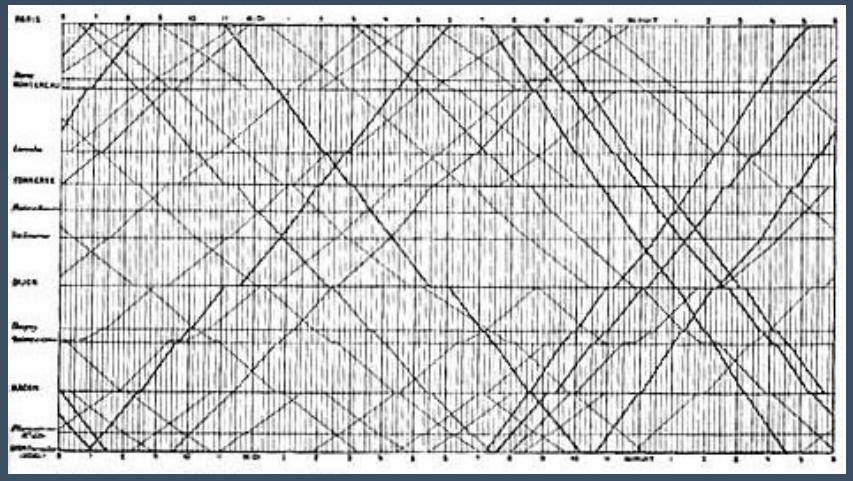
DIS		3 cls. a.m.	1 ex. a.m.	2m'l. p.m.	3c. pl. p.m.	2m'l. p.m.	Sunday. a.m.	a.m.		FARES.		
	Edinburgh	7.45	10.30	12.0	4.45	9.15	12 n'	9.15				
5	Currie	7.53	..	..	4.53	9.25	..	..	0/11	0/8	0/7	
10	Midcalder	8.19	..	12.34	5.3	9.43	12.34	9.46	1/8	1/3	1/04	
16	West-Calder	8.34	..	..	5.19	..	..	..	2/6	1/11	1/7	
21	Auchengray	8.49	..	..	5.40	..	..	..	3/8	2/8	2/3	
26	Carnwath	9.1	..	..	6.1	..	..	..	4/4	3/3	2/8	
27	Carstairs	9.10	11.15	1.22	6.10	10.25	1.22	10.32	4/7	3/6	2/10	
32	Thankerton	9.25	..	..	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	..	..	..	11/1	8/11	7/5	
74	Lockerbie	11.21	12.40	3.20	8.21	12.15	3.20	12.18	12/4	9/3	7/9	
80	Ecclefechan	11.17	..	3.35	8.37	12.31	3.35	12.33	13/40	10/0	8/4	
83	Kirtlebridge	11.46	..	..	8.46	..	..	..	13/1	10/5	8/8	
87	Kirkpatrick	12.27	..	..	8.57	..	..	..	14/6	10/11	9/1	
92	Gretna	12.29	..	4.4	9.10	12.57	4.2	12.57	15/3	11/6	9/6	
100	Annan	1.35	..	5.0	..	..	..	..	16/8	12/6	10/5	
115	Dumfries	12.22	..	5.50	..	..	..	..	16/9	15/9	12/0	
96	Rockcliffe	12.25	..	..	9.22	..	..	..	16/0	12/0	10/0	
100	Carlisle	12.30	1.22	4.18	9.30	1.14	4.18	1.15	16/8	12/6	10/5	

EDINBURGH (via Hawick Railway) TO DALKEITH.

DIS		a.m.	a.m.	noon	p.m.	p.m.	p.m.	p.m.		FARES.		
	Dalkeith	8.45	10.0	12.0	2.0	3.45	6.15	8.30				
3	Niddrie	8.53	10.8	12.8	2.8	3.53	6.23	8.38	0/4	0/30	0/2	
5	Portobello	9.0	10.15	12.15	2.15	4.0	6.30	8.45	0/6	0/7	0/0	
8	Edinburgh	9.10	10.25	12.25	2.25	4.10	6.40	8.53	0/8	0/6	0/5	

EDINBURGH (via Station at Scott Monument) TO DUNDEE.

DIS		4 cls. a.m.	3 cls. a.m.	2 cls. p.m.	2 cls. p.m.	2 cl.s p.m.	3 cls. p.m.	Sunday. a.m.	p.m.		FARES.	
	Edinburgh	6.36	9.45	12.30	3.0	4.38	7.0	7.30	4.30			
—	Leith	6.20	9.35	12.20	2.50	4.20	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	7.23	10.46	1.25	3.50	5.25	7.55	8.25	5.23	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	2/9	1/8	
16	Dysart	7.56	11.5	..	4.14	5.50	8.20	8.45	5.45	2/4	1/10	
19	Thornton Junction	8.0	11.15	..	4.22	5.58	8.28				3/0	2/10
21	Markinch (Leslie)	8.0	11.24	1.50	4.33	6.8	8.36	9.0	6.0	4/3	3/6	
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	6.28	8.58	9.21	6.29	5/8	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	8.52	12.3	..	5.13	6.51	9.21	9.41	6.41	7/3	5/10	
39	Leuchars	9.5	12.16	2.36	5.25	7.3	9.33	9.51	6.51	8/0	6/8	
—	44 St Andrews—coach	..	1.16	3.40	..	7.48	..	..	..	9/0	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	6/9	
45	Broughty Ferry	9.40	12.50	..	6.0	7.35	..	10.20	7.25	8/0	6/9	
47	Dundee	10.0	1.10	3.5	6.20	7.55	10.5	10.45	7.45	8/6	6/9	
118	Aberdeen	1.50	5.15	7.31	7.45	..	..	8.25	..	29/6	18/9	



E J Marey, Paris to Lyon Train Schedule, 1878

Possible Benefits

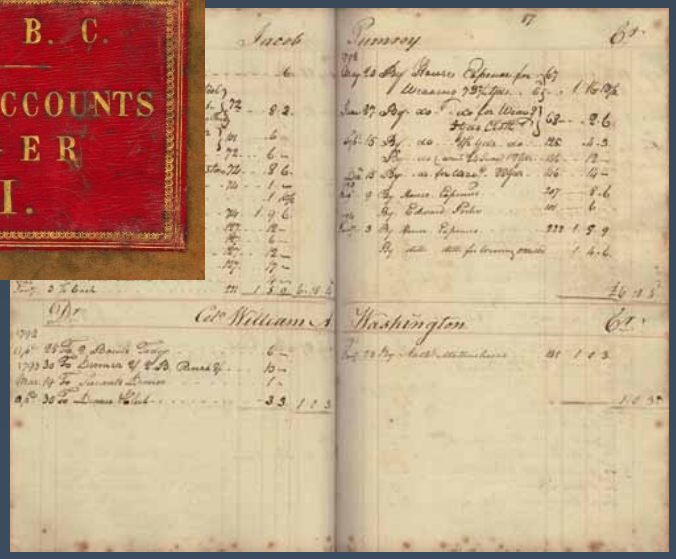
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents – Proactive Computing
- Awareness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

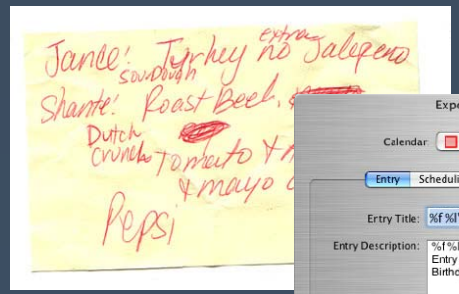
Possible Benefits

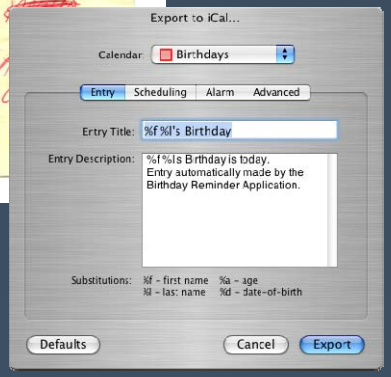
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment
- Agents – Proactive Computing
- Awareness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

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

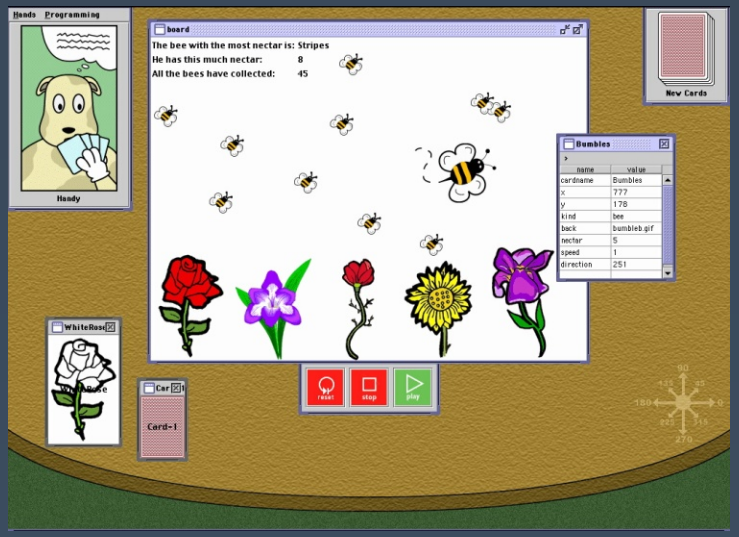






Possible Benefits

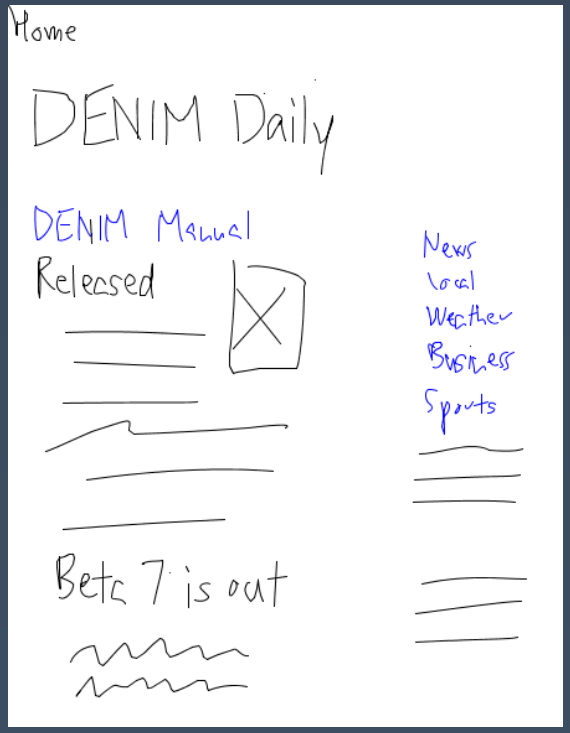
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition



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

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



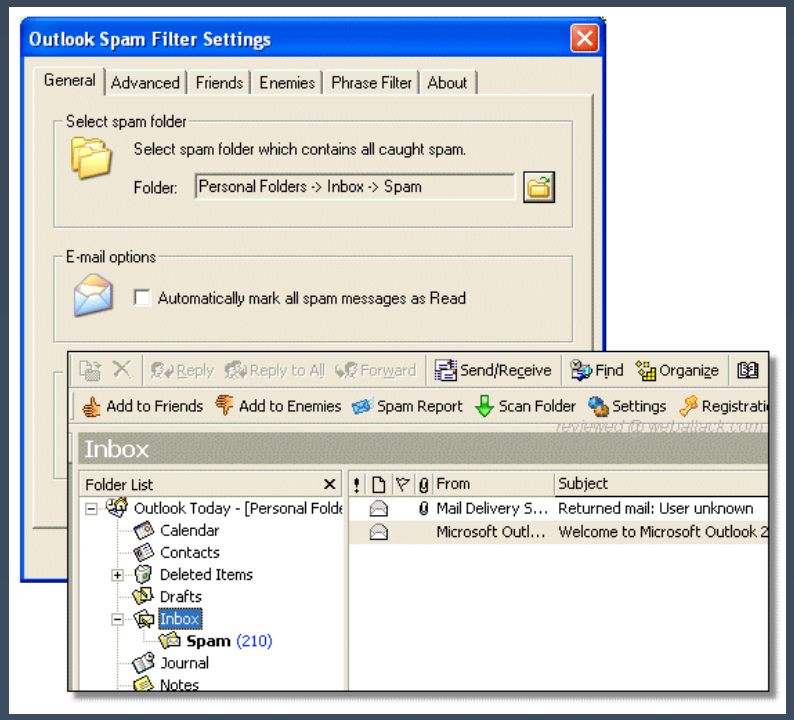
Mark Newman et al, DENIM

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

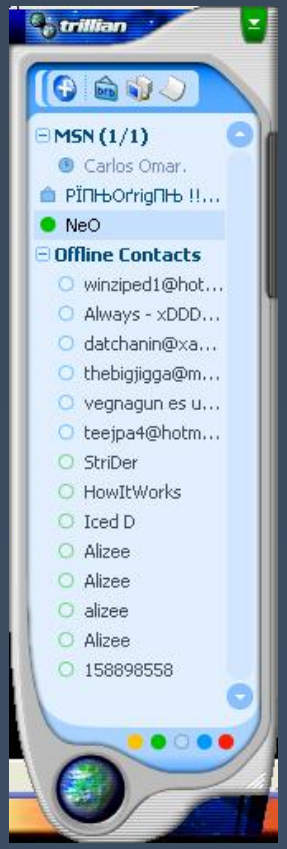
- Efficiency
- More User Control
- Affords Collaboration
- Capture and Access
- See, Feel, Sense More
- Search, Browse, Organize
- Communication
- Distributing Cognition
- Ease of use/Ease of learning
- Informality of Interaction
- Better Mappings/Embodiment



- 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 Interaction
- Better Mappings/Embodiment
- Agents – Proactive Computing
- Awareness
- Safety

Statistical analysis

- Compute aggregate statistics for each condition
  - Usually mean and standard deviation
- Compute significance (p value)
  - Likelihood that results are due to chance variation
  - $p = 0.05$ usually considered significant

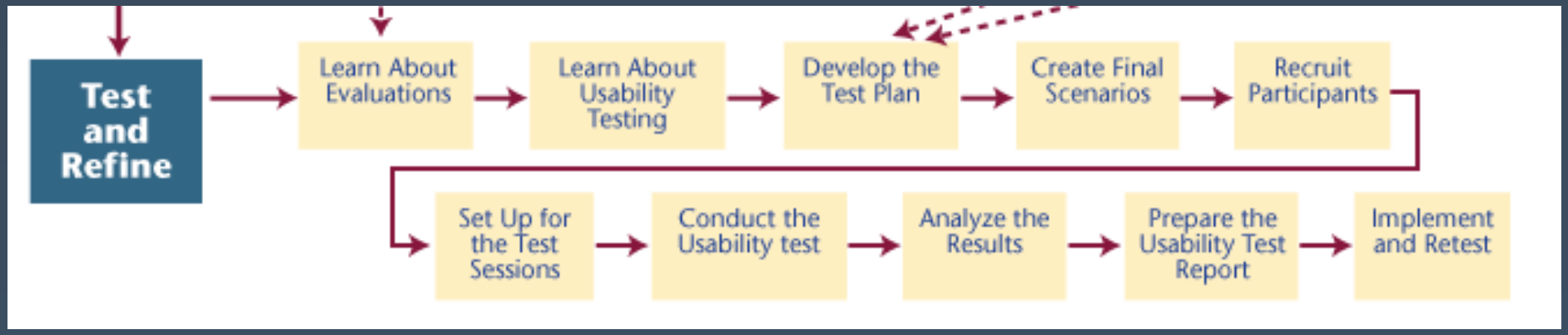
Statistical tests

- T-tests (compare 2 conditions)
- ANOVA (compare >2 conditions)
- Correlation and regression
- Many others

Manipulation of Variables

- Direct intervention: Intervening in the environment to produce the desired values
- Indirect intervention: Inducing the desired values by indirect means
- Selection: Finding cases with values and allocating them to conditions

Step-by-step testing guide



Source: Usability.gov, http://www.usability.gov/process.html

Example: Group Dynamics

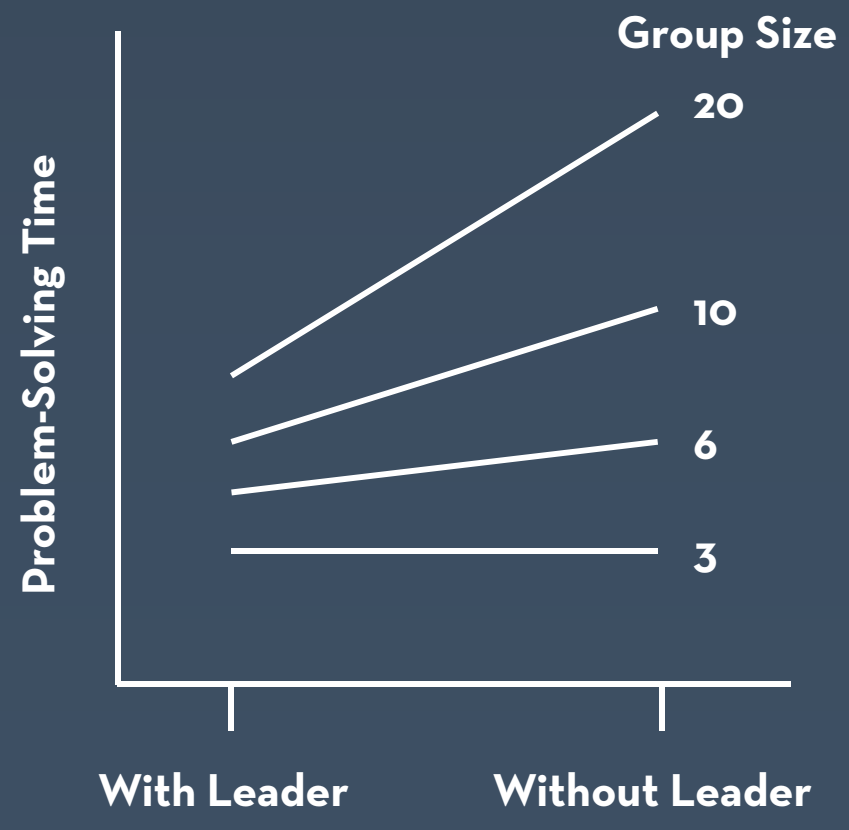
- How do leadership, group size and group affinity affect problem-solving ability?
  - Leadership (strong or not)
  - Group size (3, 6, 10, 20)
  - Group affinity (high, medium, or low)
- We have three independent variables

Number of conditions

- To isolate effect of each independent variable we consider all combinations (factorial design)
  - Strong Leader Size=3 High Affinity
  - Strong Leader Size=3 Medium Affinity
  - Strong Leader Size=3 Low Affinity
  - Strong Leader Size=6 High Affinity
  - Strong Leader Size=6 Medium Affinity
  - Strong Leader Size=6 Low Affinity
  - Strong Leader Size=10 High Affinity
  - ...
- (2 types of leadership) * (4 group sizes) * (3 affinities) = 24 combinations!

Reducing number of conditions

- Vary only some independent variables leaving others fixed (or randomly varying)
  - Will miss effects of interactions



[from Martin 04]