



# Human-Computer Interaction

An Empirical Research Perspective

**MK**  
MORGAN KAUFMANN

**I. Scott MacKenzie**

## Chapter 4 Scientific Foundations

# The Next Weeks

## Schedule:

- November 4: lecture on design of experiments
- November 6: review for midterm
  
- November 11: midterm, closed notes and books
- November 13: no lecture – get ready for in-class project beta-testing
  
- November 18: in-class project beta testing (10 minutes per group)
- November 20: in-class project beta testing continued
  
- November 25: midterm returned and discussed
- November 27: projects due, lecture on design of experiments  
user testing of project begins

# What is Research?

- “Research” means different things to different people
- Often just a word adding weight to an assertion (“*Our research shows that...*”)
- ISP television ad:<sup>1</sup>

“Independent research proves our Internet service is the fastest and most reliable – period.”

- Hmm... Is this research available for public scrutiny?
- What about the independence of the research?
- *Research* has at least three definitions (next slide)

<sup>1</sup> Aired in Ontario, winter of 2008/2009 (ad for Rogers Communications Inc.).

# Research – Definition #1

- Research is...

Careful or diligent search

- Examples
  - Searching one's garden for weeds
  - Searching a computer to find all files modified on a certain date

# Research – Definition #2

- Research is...

Collecting information about a particular subject

- Examples
  - Survey voters to collect information on political opinions in advance of an election
  - Observe people using computers and collect information, such as the number of times they
    - Consulted the manual
    - Clicked the wrong button
    - Retried an operation
    - Uttered an expletive

# Research – Definition #3

- Research is...

Investigation or experimentation aimed at the discovery and interpretation of facts, the revision of accepted theories or laws in light of new facts.

- Example
  - Design and conduct a user study to test whether a new interaction technique improves on an existing interaction technique (that's what this book is all about)

# Experimentation

- A central activity in HCI research
- An experiment is sometimes called a *user study*
- Formal, standardized methodology preferred
  - Brings consistency to a body of work
  - Facilitates reviews and comparisons between different user studies

# Facts, Theories, Laws

- Facts
  - Building blocks of evidence
  - Evidence is tested to confirm hypotheses (more later)
- Theory
  - An hypothesis assumed for the sake of argument
  - A scientifically accepted body of principles that explain phenomena
- Law
  - More constraining, more formal, more binding
  - A relationship that is invariable under given conditions
  - HCI involves humans, so laws are of questionable value

# Law vs. Theory

- A law tells you WHAT happens to something. It won't really be proven wrong.
- A theory tells you WHY something happens to something. It is not really disputed, but it has the potential to be proven wrong. It's like a summary of the results of many, many tested hypotheses.

# Examples

- Law
  - Newton's Law of Gravity; if something is dropped in a place with gravity, it falls.
- Theory
  - The theory of relativity: that the laws of physics are the same for all observers
  - The theory of evolution by natural selection: that the observed changes in species occur due to selection of well adapted specimens over less well adapted specimens.
  - Big Bang theory: that the universe began as an infinitely small point that underwent expansion to form the universe as we know it today.

# Fitts' Law<sup>1</sup>

- HCI's best known “law”
- Fitts proposed a model, not a law
- Fitts' law is a behavioural, predictive, and descriptive model of human motor behaviour
- It is a “law” only in that other researchers took up the label as a celebration of the robustness and importance of Fitts' work
- See **HCI:ERP** (Chapter 7)

<sup>1</sup> Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391.

Let's look at some other characteristics of research that are not encompassed in the definitions...

# Research Must Be Published

- Publication is the final step
- Also an essential step
- *Publish or perish!*
  - Edict for researchers in all fields, and particularly in academia
- Until it is published, research cannot achieve its critical goal:
  - Extend, refine, or revise the existing body of knowledge in the field

# Peer Review

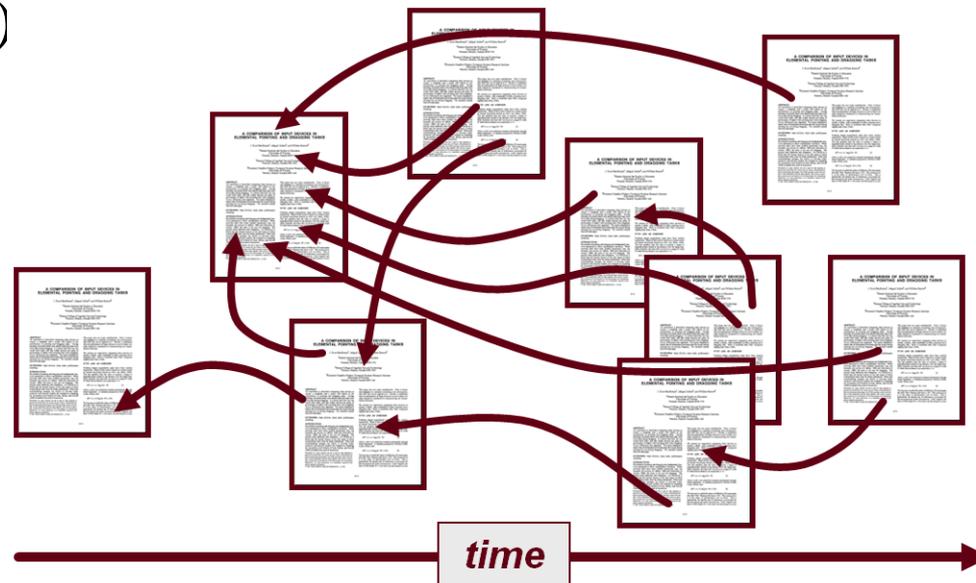
- Research submitted for publication is reviewed by peers – other researchers doing similar research
- Only research meeting a high standard of scrutiny is accepted for publication
  - Are the results novel and useful?
  - Does the evidence support the conclusions?
  - Does the methodology meet the expected standards for the field?
- Accepted research is published and archived
- The final step is complete

# Patents

- Some research develops into bona fide inventions
- A researcher/company may wish to maintain ownership of (profit from) the invention
- Patenting is an option
- The patent application describes
  - Previous related work
  - How the invention addresses a need
  - The best mode of implementation
- If the application is granted, the patent is issued
- Note: A patent is a publication; thus patenting meets the must-publish criterion for research

# Citations, References, Impact

- Citations, like hyperlinks, connect research to other research
- Through citations, a body of research takes shape
- The number of citations to a research paper is an indication of the paper's impact
- Can you spot the high-impact paper below? (arrows are citations)



# Google Scholar – H-index<sup>1</sup>

- Since the arrival of Google Scholar, citation counts are easy to gather
- Can be gathered for papers, journals, etc.
- Can also be gathered for researchers
- H-index is a measure of the impact of a researcher
- If a researcher's publications are ranked by the number of citations, the H-index is the point where the rank equals the number of citations; i.e., a researcher with H-index =  $n$  has  $n$  publications each with  $n$  or more citations
- A respectable H-index (although debatable) is “number of years since PhD”

<sup>1</sup> Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16568-16572. 17

# Research Must Be Reproducible

- Research that cannot be replicated is useless
- A high standard or reproducibility is essential
- The research write-up must be sufficiently detailed to allow a skilled researcher to replicate the research if he/she desired
- The easiest way to ensure reproducibility is to follow a standardized methodology
- Many great advances in science pertain to methodology (e.g., Louis Pasteur's detailed disclosure of the methodology used in his research in microbiology)
- The most cited research paper is a “method paper”<sup>1</sup> (see Google Scholar for the latest citation count)

<sup>1</sup> Lowry, O. H., Rosenbrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry*, 193, 265-275.

# Research vs. Engineering vs. Design

- Researchers often work closely with engineers and designers, but the skills each brings are different
- Engineers and designers are in the business of building things, bringing together the best in *form* (design emphasis) and *function* (engineering emphasis)
- One can image that there is a certain tension, even trade-off, between form and function
- Sometimes, things don't go quite as planned →

# Form Trumpeting Function

- The photo below shows part of a laptop computer
- The form is elegant – smooth, shiny, metallic
- The touchpad design (or is in engineering?) has a problem
- No tactile sense at the sides of the touchpad



# Duct Tape To The Rescue



**A true story!**

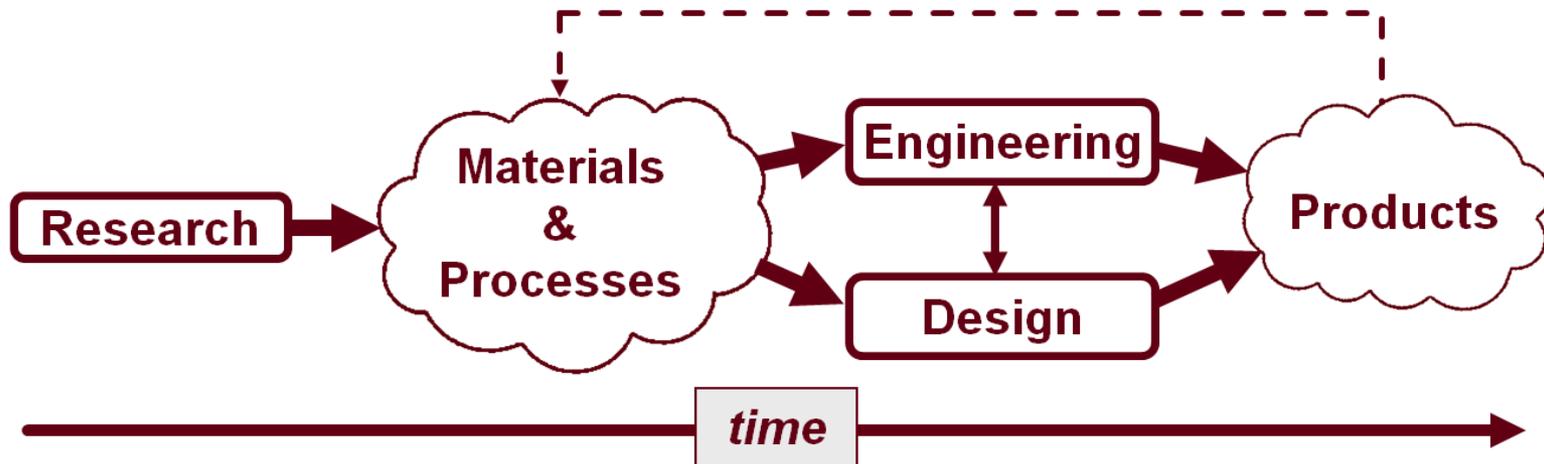
# Function Trumpeting Form?



# Research Milieu

- Engineering and design are about products
- Research is not about products
- Research is narrowly focused
- Research questions are small in scope
- Research is incremental, not monumental
- Research ideas build on previous research ideas
- Good ideas are refined, advanced (into new ideas)
- Bad ideas are discarded, modified
- Products come later, much later

# Schematic



# Example: Apple *iPhone* (2007)

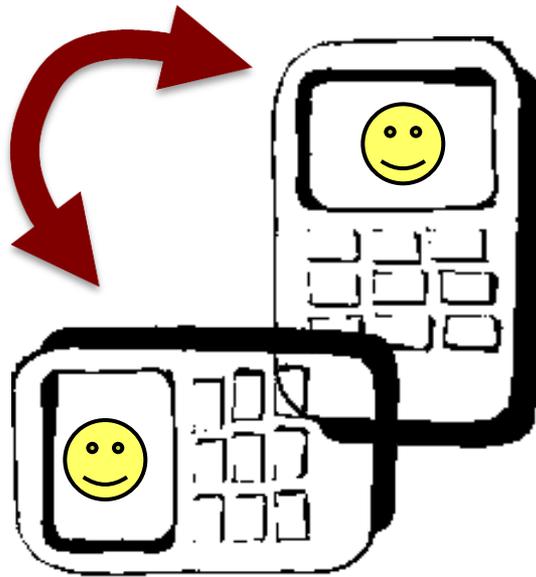


## *iPhone* Gestures:

- Tilt
- Multitouch
- Flick

# Tilt

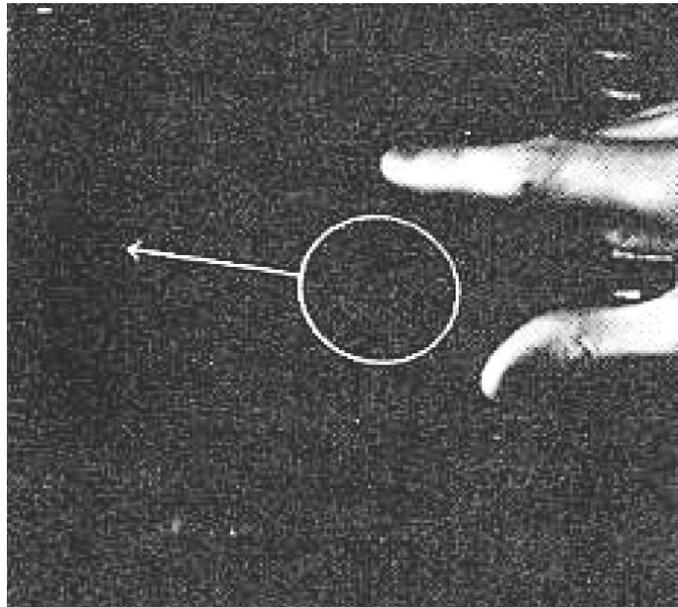
- Research on tilt as an interaction primitive dates at least to 1998<sup>1</sup>



<sup>1</sup> Harrison, B., Fishkin, K. P., Gujar, A., Mochon, C., & Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. *Proc CHI '98*, 17-24, New York: ACM.

# Multitouch

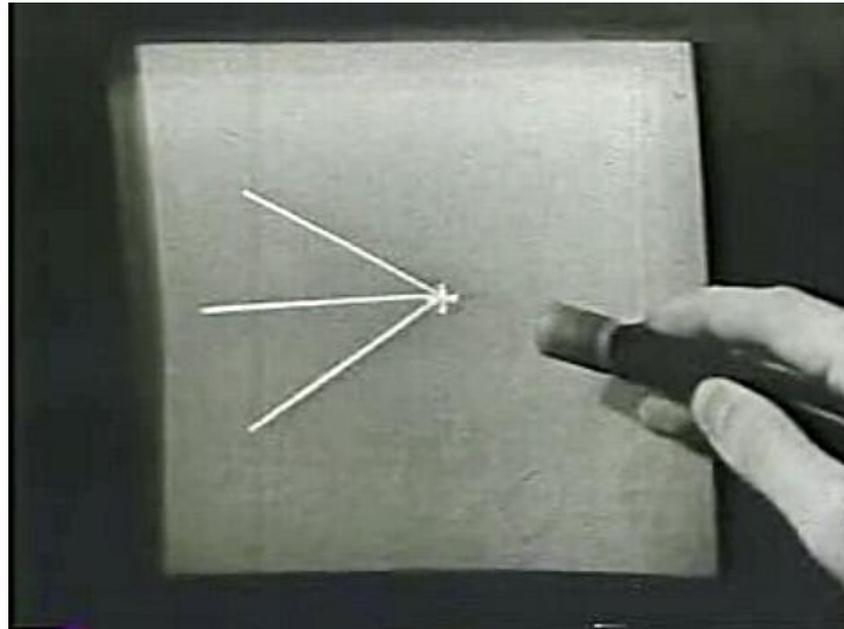
- Research on multitouch as an interaction primitive dates at least to 1978<sup>1</sup>



<sup>1</sup> Herot, C. F., & Weinzapfel, G. (1978). One-point touch input of vector information for computer displays. *Proceedings of SIGGRAPH 1978*, 210-216, New York: ACM.

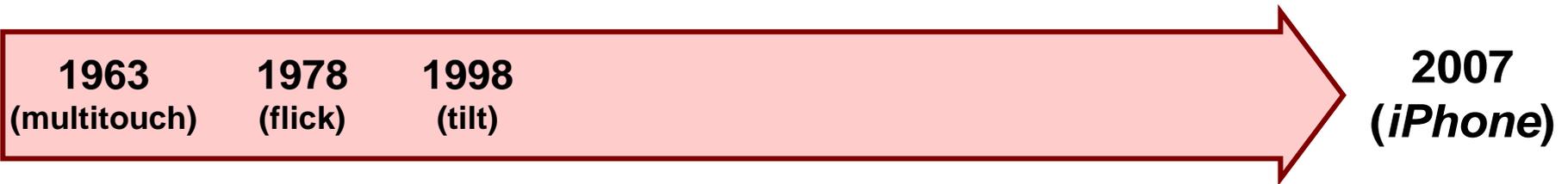
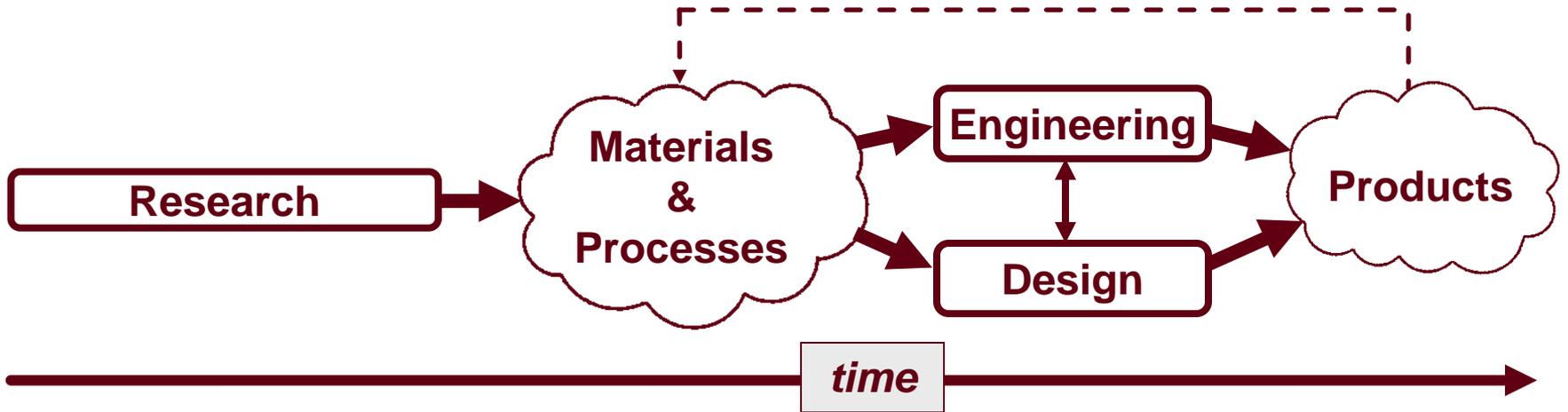
# Flick

- Research on flick as an interaction primitive dates at least to 1963<sup>1</sup>



<sup>1</sup> Sutherland, I. E. (1963). Sketchpad: A man-machine graphical communication system. *Proceedings of the AFIPS Spring Joint Computer Conference*, 329-346, New York: ACM.

# So...



# Design as Research

- Gaver and Bowers opine on the struggle for designers to also be researchers:<sup>1</sup> (paraphrased)
  - *Do we need to add research questions or methodological rigour to design practice for it to count as research?*
  - *Do we have to change design practices to make our contributions to HCI look more like research?*
  - *Is the result still design, or have we lost something in the process?*
  - *These questions have been vexing the HCI design community – and us – for some time. The problem is that novel products alone do not seem sufficient to count as research.*

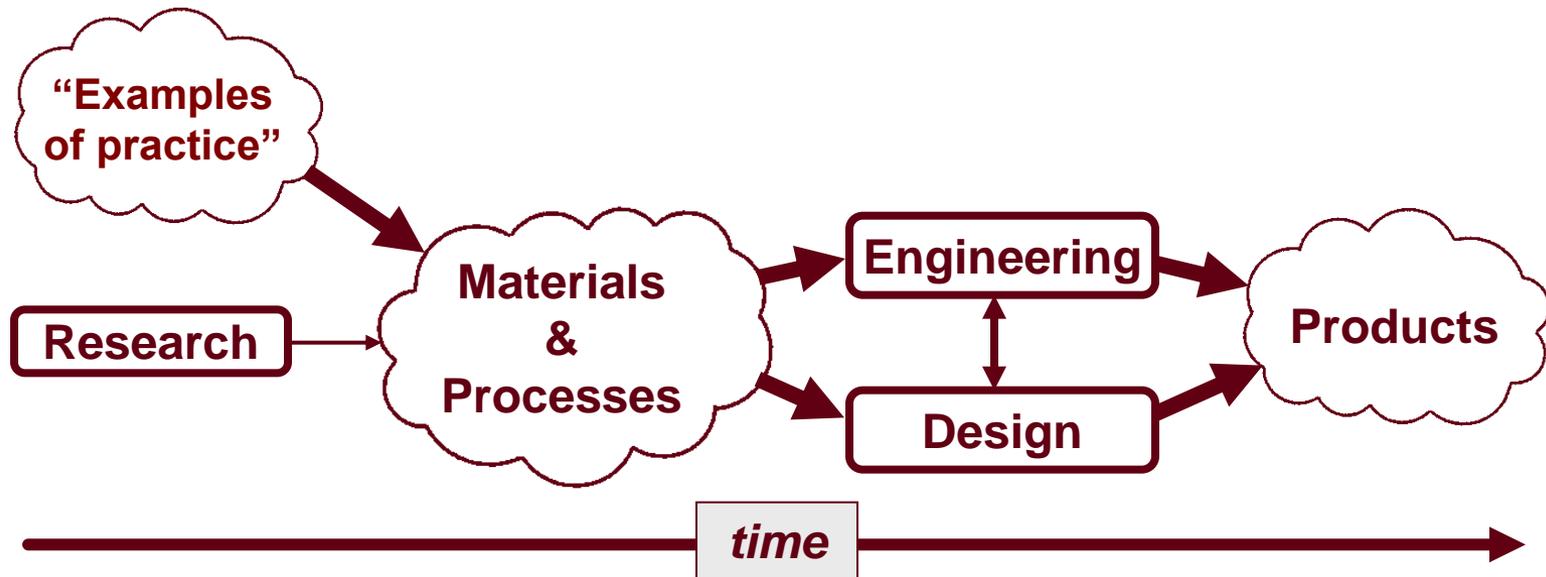


Photostroller

<sup>1</sup> Gaver, B., & Bowers, J. (2012, July/August). Annotated portfolios. *interactions*, 40-49.

# On Materials and Processes

- Gaver and Bowers also comment on the materials and processes they use in design:
  - *It was by looking at specific examples of practice that we found guidance for our work.*



# “Empirical” Research

- Empirical:
  - Originating in or based on observation or experience
  - Relying on experience or observation alone without due regard for system or theory (i.e., don’t be blinded by pre-conceptions)
- Example: Nicolas Copernicus (1473-1543)
  - Prevailing system or theory: celestial bodies revolved around the earth
  - Copernicus made astronomical observations that cut against this view
  - Result: heliocentric cosmology (the earth and planets revolve around the sun)

# “Empirical” Research (2)

- Empirical: (by another definition)
  - Capable of being verified or disproved by observation or experiment
- HCI research
  - Framed by hypotheses
  - Methodology to test hypotheses
  - Experiments (aka user studies) are the vehicle
  - Hypotheses must be sufficiently narrow and clear to allow for verification or disproof (more on this later; see Research Questions)

# Research Methods

- Observational method
- Experimental method
- Correlational method

# Observational Method

- Example methods:
  - Interviews, field investigations, contextual inquiries, case studies, field studies, focus groups, think aloud protocols, story telling, walkthroughs, cultural probes, etc.
- Focus on qualitative assessments (cf. quantitative)
- Relevance vs. precision
  - High in relevance (behaviours studied in a natural setting)
  - Low in precision (lacks control available in a laboratory)
- Goal: discover and explain reasons underlying human behaviour (*why* or *how*, as opposed to *what*, *where*, or *when*)

# Experimental Method

- Aka *scientific method*
- Controlled experiments conducted in lab setting
- Relevance vs. precision
  - Low in *relevance* (artificial environment)
  - High in *precision* (extraneous behaviours easy to control)
- At least two variables:
  - *Manipulated variable* (aka *independent variable*)
  - *Response variable* (aka *dependent variable*)
- Cause-and-effect conclusions possible (changes in the manipulated variable *caused* changes in the response variable)

# Correlational Method

- Look for relationships between variables
- Observations made, data collected
  - Example: are user's privacy settings while social networking related to their age, gender, level of education, employment status, income, etc.
- Non-experimental
  - Interviews, on-line surveys, questionnaires, etc.
- Balance between relevance and precision (some quantification, observations not in lab)
- Cause-and-effect conclusions not possible

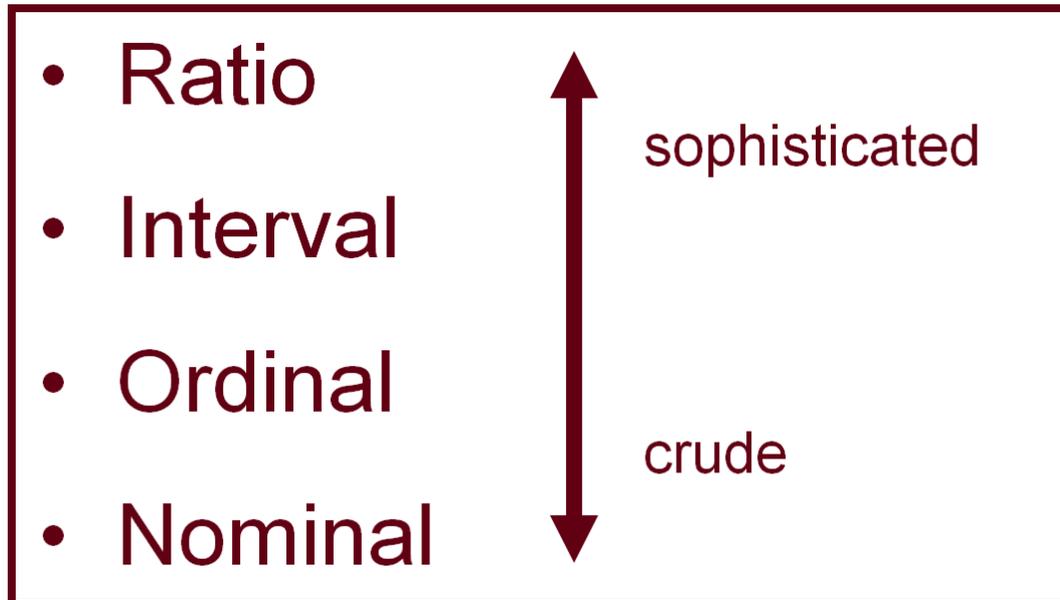
# Correlation is Not Causation

- As ice cream sales increase, the rate of drowning deaths increases sharply.
  - therefore, ice cream consumption causes drowning
  - true?
- Sleeping with one's shoes on is strongly correlated with waking up with a headache.
  - therefore, sleeping with one's shoes on causes headache
  - true?

# Observe and Measure

- Foundation of empirical research
- Observation is the starting point; observations are made...
  - By the apparatus
  - By a human observer
- Manual observation
  - Log sheet, notebooks
  - Screen capture, photographs, videos, etc.
- Measurement
  - With measurement, anecdotes (*April showers bring May flowers*) turn to empirical evidence
  - “*When you cannot measure, your knowledge is of a meager and unsatisfactory kind*” (Kelvin)

# Scales of Measurement



# Nominal Data

- Nominal data (aka *categorical data*) are arbitrary codes assigned to attributes; e.g.,
  - 1 = male, 2 = female
  - 1 = mouse, 2 = touchpad, 3 = pointing stick
- The code needn't be a number; i.e.,
  - M = male, F = female
- Obviously, the statistical mean cannot be computed on nominal data
- Usually it is the count that is important
  - “Are females or males more likely to...”
  - “Do left handers or right handers have more difficulty with...”
  - Note: The count itself is a ratio-scale measurement
- (example on next slide)

# Nominal Data – HCI Example

- Task: Observe students “on the move” on university campus
- Code and count students by...
  - Gender (male, female)
  - Mobile phone usage (not using, using)

| Gender | Mobile Phone Usage |       | Total | %     |
|--------|--------------------|-------|-------|-------|
|        | Not Using          | Using |       |       |
| Male   | 683                | 98    | 781   | 51.1% |
| Female | 644                | 102   | 746   | 48.9% |
| Total  | 1327               | 200   | 1527  |       |
| %      | 86.9%              | 13.1% |       |       |

**Real Data!**

# Ordinal Data

- Ordinal data associate an order or rank to an attribute
- The attribute is any characteristic or circumstance of interest; e.g.,
  - Users try three GPS systems for a period of time, then rank them: 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> choice
- More sophisticated than nominal data
  - Comparisons of “greater than” or “less than” possible
- (example on next slide)

# Ordinal Data – HCI Example

How many email messages do you receive each day?

1. None (I don't use email)
2. 1-5 per day
3. 6-25 per day
4. 26-100 per day
5. More than 100 per day

# Interval Data

- Equal distances between adjacent values
- But, no absolute zero
- Classic example: temperature ( $^{\circ}\text{F}$ ,  $^{\circ}\text{C}$ )
- Statistical mean possible
  - E.g., the mean midday temperature during July
- Ratios not possible
  - Cannot say  $10^{\circ}\text{C}$  is twice  $5^{\circ}\text{C}$

# Interval Data – HCI Example

- Questionnaires often solicit a level of agreement to a statement
- Responses on a Likert scale
- Likert scale characteristics:
  1. Statement soliciting level of agreement
  2. Responses are symmetric about a neutral middle value
  3. Gradations between responses are equal (more-or-less)
- Assuming “equal gradations”, the statistical mean is valid (and related statistical tests are possible)
- Likert scale example → (next slide)

# Interval Data – HCI Example (2)

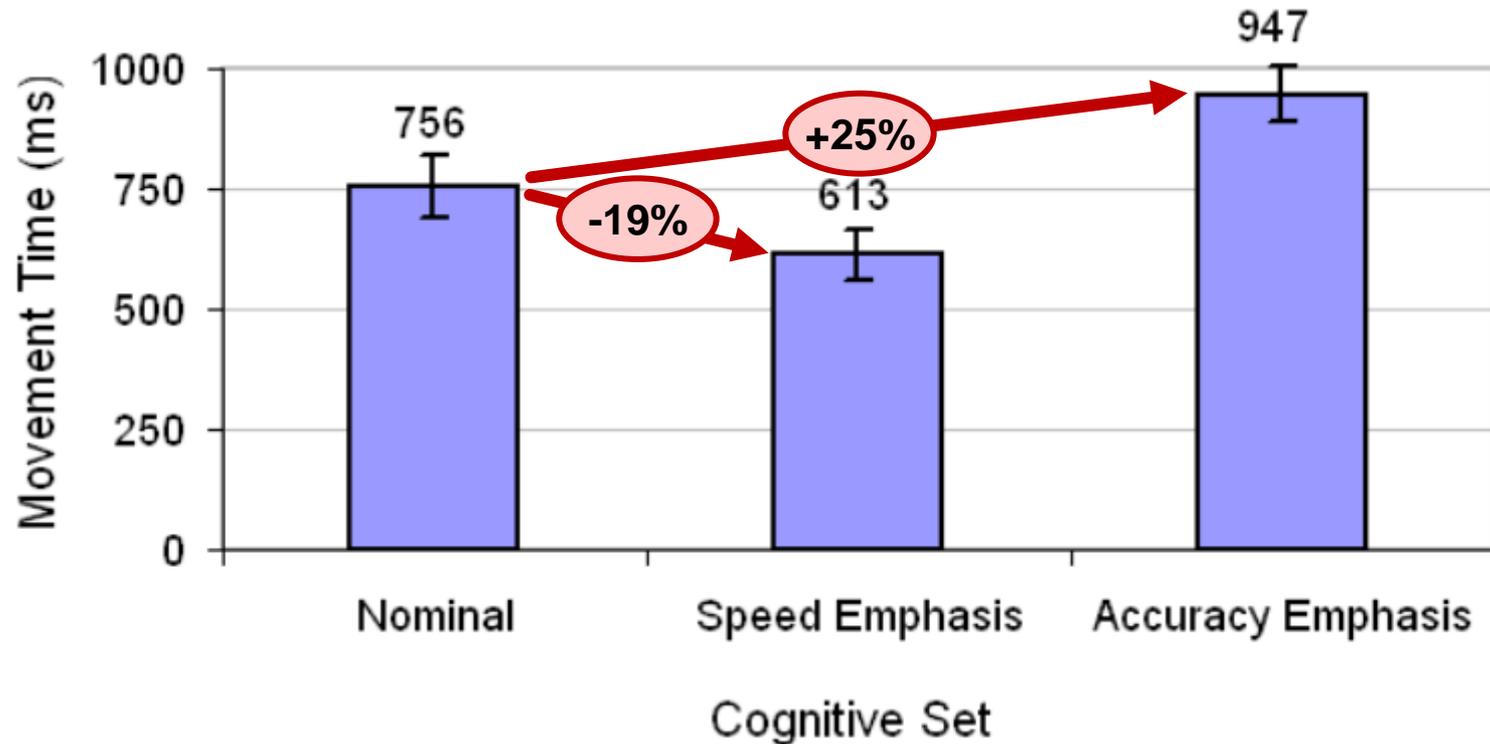
Please indicate your level of agreement with the following statements.

|   | Strongly disagree | Mildly disagree | Neutral | Mildly agree | Strongly agree |
|---|-------------------|-----------------|---------|--------------|----------------|
| It is safe to talk on a mobile phone while driving.                   | 1                 | 2               | 3       | 4            | 5              |
| It is safe to read a text message on a mobile phone while driving.    | 1                 | 2               | 3       | 4            | 5              |
| It is safe to compose a text message on a mobile phone while driving. | 1                 | 2               | 3       | 4            | 5              |

# Ratio Data

- Most sophisticated of the four scales of measurement
- Preferred scale of measurement
- Absolute zero, therefore many calculations possible
- Summaries and comparisons are strengthened
- A “count” is a ratio-scale measurement
  - E.g., “time” (the number of seconds to complete a task)
- Enhance counts by adding further ratios where possible
  - Facilitates comparisons
  - Example – a 10-word phrase was entered in 30 seconds
    - Bad:  $t = 30$  seconds
    - Good: Entry rate =  $10 / 0.5 = 20$  wpm

# Ratio Data – HCI Example<sup>1</sup>



$$F_{2,34} = 372.7, p < .0001$$

<sup>1</sup> MacKenzie, I. S., & Isokoski, P. (2008). Fitts' throughput and the speed-accuracy tradeoff. *Proc CHI 2008*, 1633-1636, New York: ACM.

# Research Questions

- We conduct empirical research to answer (and raise!) questions about UI designs or interaction techniques
- Consider the following questions:
  - Is it viable?
  - Is it better than current practice?
  - Which design alternative is best?
  - What are the performance limits?
  - What are the weaknesses?
  - Does it work well for novices?
  - How much practice is required?

# Testable Research Questions

- Preceding questions, while unquestionably relevant, are not testable
- Try to re-cast as testable questions (even though the new question may appear less important)
- Scenario...
  - You have invented a new text entry technique for touchscreen mobile phones, and you think it's pretty good. In fact, you think it is better than the Qwerty soft keyboard (QSK). You decide to undertake a program of empirical enquiry to evaluate your invention. What are your research questions?

# Research Questions (2)

- Very weak

*Is the new technique any good?*

- Weak

*Is the new technique better than QSK?*

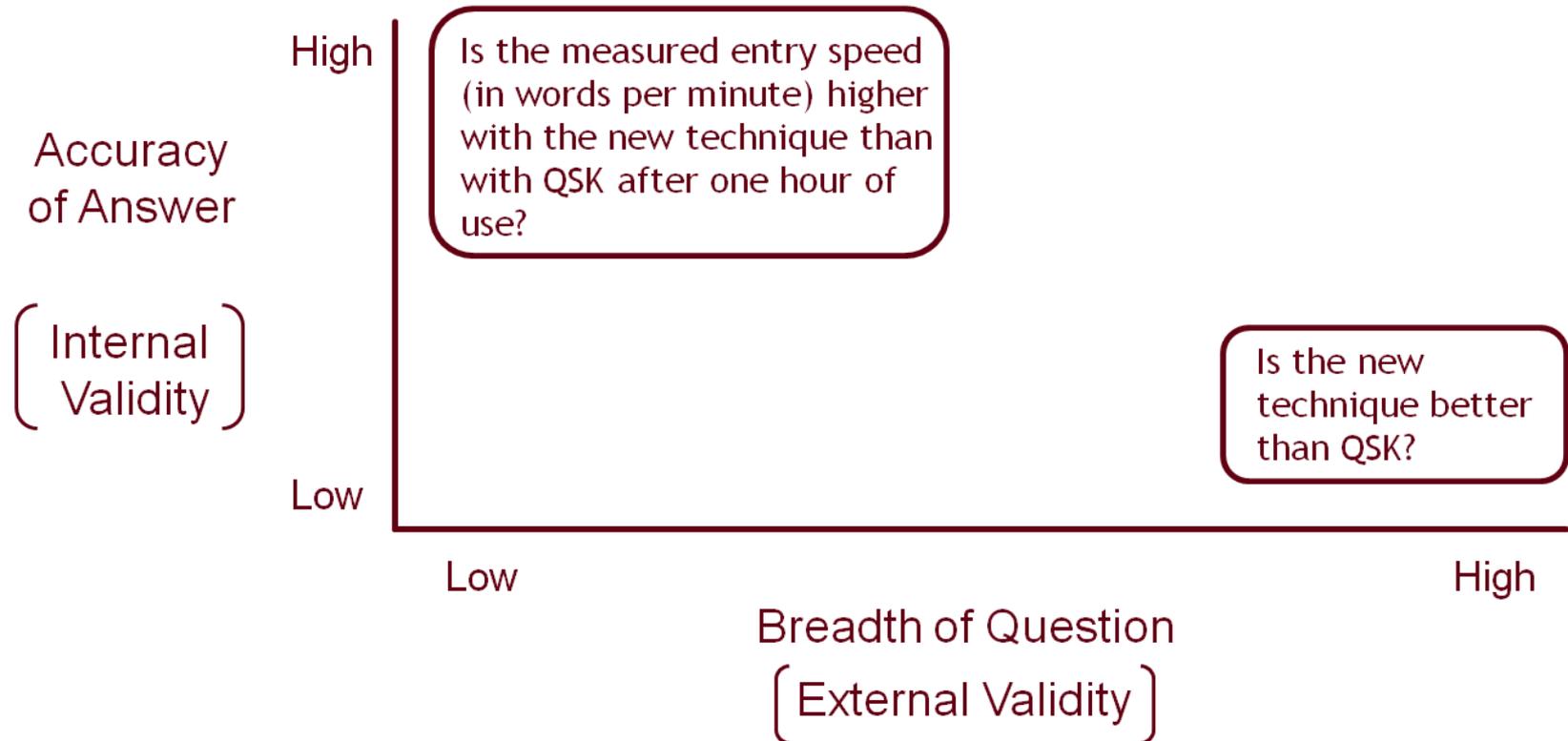
- Better

*Is the new technique faster than QSK?*

- Better still

*Is the measured entry speed (in words per minute) higher for the new technique than for QSK after one hour of use?*

# A Tradeoff



# Internal Validity

- Definition:
  - The extent to which the effects observed are due to the test conditions (e.g., multitap vs. new)
- Statistically, this means...
  - Differences (in the means) are due to inherent properties of the test conditions
  - Variances are due to participant differences (“pre-dispositions”)
  - Other potential sources of variance are controlled or exist equally or randomly across the test conditions

# External Validity

- Definition:
  - The extent to which results are generalizable to other *people* and other *situations*
- People
  - The participants are *representative* of the broader intended population of users
- Situations
  - The *test environment* and *experimental procedures* are representative of real world situations where the interface or technique will be used

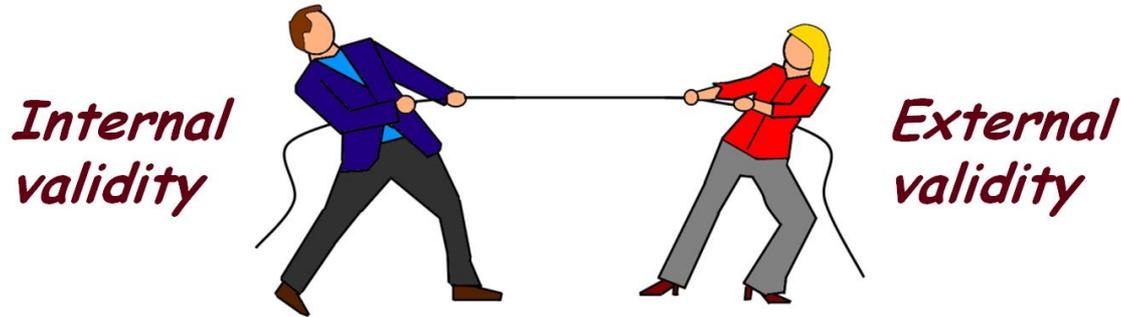
# Test Environment Example

- Scenario...
  - You wish to compare two input devices for remote pointing (e.g., at a projection screen)
- External validity is improved if the test environment mimics expected usage
- Test environment should probably...
  - Use a large display or projection screen (not a desktop monitor)
  - Position participants at a significant distance from screen (rather than close up)
  - Have participants stand (rather than sit)
  - Include an audience!
- But... is internal validity compromised?

# Experimental Procedure Example

- Scenario...
  - You wish to compare two text entry techniques for mobile devices
- External validity is improved if the experimental procedure mimics expected usage
- Test procedure should probably have participants...
  - Enter personalized paragraphs of text (e.g., a paragraph about a favorite movie)
  - Edit and correct mistakes as they normally would
- But... is internal validity compromised?

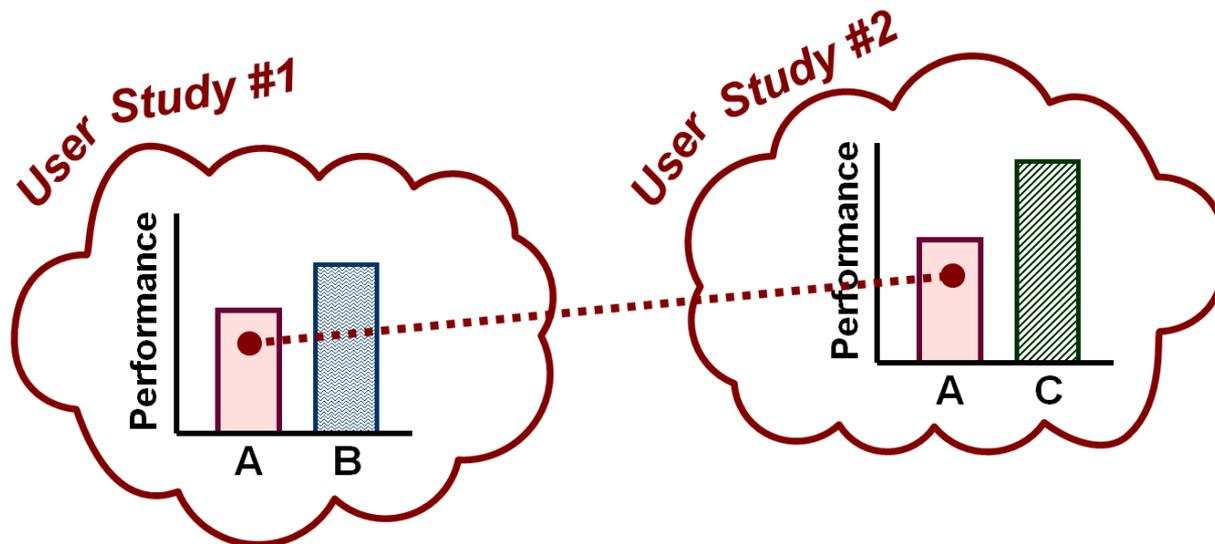
# The Tradeoff



- There is tension between internal and external validity
- The more the test environment and experimental procedures are “relaxed” (to mimic real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as pondering, distractions, fiddling, or secondary tasks

# Comparative Evaluations

- Preferable to do a comparative evaluation rather than one-of
- More insightful results obtained
- Factorial experiments require comparison, because there must be at least one independent variable with at least two levels
- If one condition is a base line; comparisons possible between studies (assuming similar methodology)



# Relationships: Circumstantial & Causal

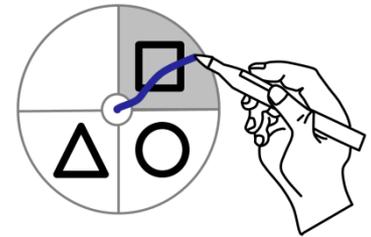
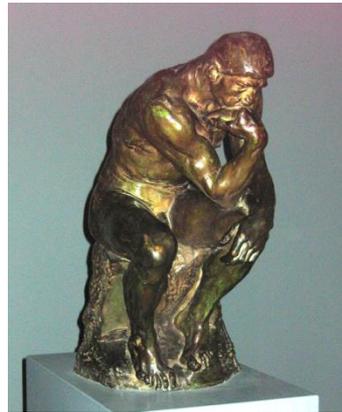
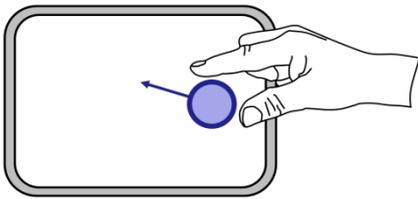
- As noted above...
  - Correlational methods → circumstantial relationships
  - Experimental methods → causal relationships
- Causal-and-effect conclusions not possible if the independent variable is a *naturally occurring attribute* of participants (e.g., gender, personality type, handedness, first language, political viewpoint)
- These attributes are legitimate independent variables
- But, they cannot be assigned to participants; hence causal relationships not valid
- Why: Cannot avoid confounding variables (Chapter 5)

# Examples: Circumstantial & Causal

- Smoking causes Cancer?
- Multitap vs. predictive input on mobile phones – predictive is faster?

# Research Topics

- Finding a research topic is a challenge (for students... and for seasoned researchers too!)
- Consult **HCI:ERP** for discussions and stories



- Four tips:
  1. Think small
  2. Replicate
  3. Know the literature
  4. Think inside the box

# Tip #1 - Think Small

- Looking for that big idea?
- Advice: Forget it (besides, it isn't necessary)
- Research questions are small, narrowly focused
- Pursue several small, related research topics and before you know it, a dissertation topic is formed

# Tip #2 - Replicate

- Seems odd: where's the research in simply replicating what was done before?
- Of course, there is no research in replication, but the trick is in the path to replicating
- Replicating prior research is a lot of work
- Along the way, you will undoubtedly discover small and novel improvements – things to try
- A little tweak here, a small modification there
- BTW, you might not find a novel idea until well into the process (perhaps afterward!)

# Tip #3 – Know The Literature

- Whatever topic interests you, read the literature
- E.g., *social networking, gaming*
- If too broad, narrow (e.g., *privacy settings* in social networking, *avatars* in gaming)
- Read papers, open a spreadsheet, tabulate variables in the methodology and the findings
- Chaotic at first, order and shape will emerge (eventually)
- With some luck (and further study) a research topics will emerge

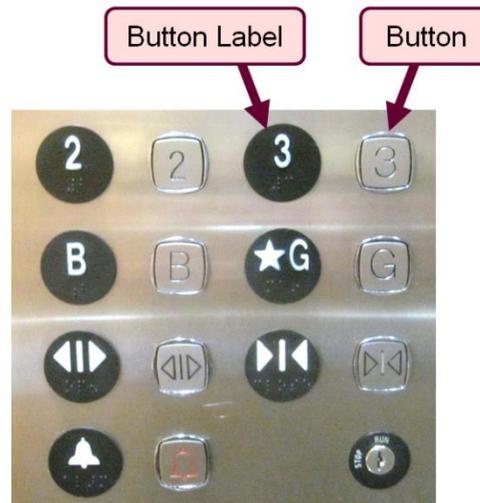
# Example Table<sup>1</sup>

| Study (1 <sup>st</sup> author)  | Number of Keys <sup>a</sup> | Direct/Indirect | Scanning | Number of Participants | Speed <sup>b</sup> (wpm) | Notes   |
|---|-----------------------------|-----------------|----------|------------------------|--------------------------|---|
| Bellman [2]   | 5                           | Indirect        | No       | 11                     | 11                       | 4 cursors keys + SELECT key. Error rates not reported. No error correction method.  |
| Dunlop [4]  | 4                           | Direct          | No       | 12                     | 8.90                     | 4 letter keys + SPACE key. Error rates reported as "very low."  |
| Dunlop [5]  | 4                           | Direct          | No       | 20                     | 12                       | 4 letter keys + 1 key for SPACE/NEXT. Error rates not reported. No error correction method.   |
| Tanaka-Ishii [25]   | 3                           | Direct          | No       | 8                      | 12+                      | 4 letters keys + 4 keys for editing, and selecting. 5 hours training. Error rates not reported. Errors corrected using CLEAR key.   |
| Gong [7]  | 3                           | Direct          | No       | 32                     | 8.01                     | 3 letter keys + two additional keys. Error rate = 2.1%. Errors corrected using DELETE key.  |
| MacKenzie [16]  | 3                           | Indirect        | No       | 10                     | 9.61                     | 2 cursor keys + SELECT key. Error rate = 2.2%. No error correction method.  |
| Baljko [1]  | 2                           | Indirect        | Yes      | 12                     | 3.08                     | 1 SELECT key + BACKSPACE key. 43 virtual keys. RC scanning. Same phrase entered 4 times. Error rate = 18.5%. Scanning interval = 750 ms.  |
| Simpson [24]  | 1                           | Indirect        | Yes      | 4                      | 4.48                     | 1 SELECT key. 26 virtual keys. RC scanning. Excluded trials with selection errors or missed selections. No error correction. Scanning interval = 525 ms at end of study.  |
| Koester [10]  | 1                           | Indirect        | Yes      | 3                      | 7.2                      | 1 SELECT key. 33 virtual keys. RC scanning with word prediction. Dictionary size not given. Virtual BACKSPACE key. 10 blocks of trials. Error rates not reported. Included trials with selection errors or missed selections. Fastest participant: 8.4 wpm. |
| <sup>a</sup> For "direct" entry, the value is the number of letter keys. For "indirect" entry, the value is the total number of keys.<br><sup>b</sup> The entry speed cited is the highest of the values reported in each source, taken from the last block if multiple blocks. |                             |                 |          |                        |                          |   |

<sup>1</sup> MacKenzie, I. S. (2009). The one-key challenge: Searching for an efficient one-key text entry method. *Proc ASSETS 2009*, 91-98, New York: ACM.

# Tip #4 – Think Inside The Box

- Think *outside* the box → dispense with accepted beliefs and assumptions (in the box), and think in a way that assumes nothing and challenges everything
- No need
- Think *inside* the box: just get on with your day; but at every juncture, every interaction, think and question
- Our everyday foibles are fertile ground for research topics



See **HCI:ERP**  
for discussion

# Tip #5 – Avoid Over-stimulus



# Thank You

