

## CHAPTER 2: HUMAN FACTOR

- Sensory stimuli and sense organs are physiological;
  - 5 senses, vision, hearing (audition), touch (tactition), smell & taste
- Responders: motor control; limbs, voice, eyes, brain & perception
  - lengths of lines show area of motor cortex
- Better performance is typically associated with faster or more accurate behavior, and this leads to a fundamental property of human performance, the speed-accuracy trade-off: go faster and errors increase; slow down and accuracy improves.

Variable	Advantage	Disadvantage
Random	Improves external validity by using a variety of situations and people.	Compromises internal validity by introducing additional variability in the measured behaviours.
Control	Improves internal validity since variability due to a controlled circumstance is eliminated	Compromises external validity by limiting responses to specific situations and people.

## CHAPTER 4: SCIENTIFIC FOUNDATIONS

- Research: investigation/experimentation aimed at discovery & interpretation of facts
  - And revision of accepted theories/laws in light of new facts
- A hypothesis is a predictive statement about the relationship between 2 or more variables. Research questions are similar to hypotheses, but they are in question format.
- Engineers & designers bring together form (design) & function (engineering)
- Empirical research: can be verified/disproved by observation/experiment
- PUBLISH OR PERISH - research must be reproducible
- Observational method: HCI usability evaluation, high relevance, low precision
- Experimental method: HCI user study, low relevance, high precision -> causal relationships
- Correlational method: looks for relationships between variables -> circumstantial relationships
- Measurement scales: nominal, ordinal, interval, ratio
  - Nominal aka categorical data - ordinal data is order/rank ie greater/less than
  - Interval data have equal distances between adjacent values, no absolute zero, ratio !possible ie temp (F, C)
  - Ratio data, have absolute zero, support many calculations to summarize, compare and test data
- Internal validity: extent to which effects observed are due to test conditions
  - Statistically, difs in means are due to inherent properties of test conditions, variances are due to participant difs ("pre-dispositions"), other variance are controlled or exist equally or randomly across test conditions
- External validity: extent to which experimental results are generalizable to other people/situations
  - People: participants representative of intended population
  - Situations: test environment and experimental procedures representative of real world where technique is used
  - Improved if experimental procedure mimics expected usage
    - More test and procedures mimic the real-world aka released, then results are uncontrollable
- Causal aka cause-and-effect relationship - Circumstantial relationship (not causal)

2 x 2

A	B
B	A

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

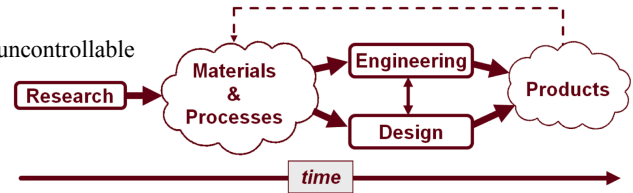
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

5 x 5

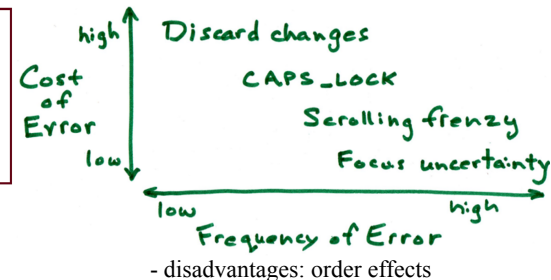
A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

## CHAPTER 5: DESIGNING HCI EXPERIMENTS

- Methodology: way an experiment is designed and carried out
  - Allen Newell "Science is method. Everything else is commentary"
- Critical for ethics approval: research methodology, risks/benefits, right not to participate, right to anonymity/confidentiality
- Independent v.: circumstance/characteristic is manipulated in experiment to elicit change in response while interacting with comp
  - Must have at least 2 levels (aka different things to test)
- Dependent v.: any observable, measured human behavior, depends on what participant does
  - Must be clearly defined
- Control v.: circumstance (not under investigation) kept constant while testing effect of an IV
- Random v.: circumstance allowed to vary randomly (more variability introduced in measures (bad) but results generalizable (good))
- Confounding v.: circumstance varies systematically with IV
- Experiment task must "elicit a change"
  - good task represent, discriminate
- Procedure: encompasses everything occurs with participants
- # participants
  - too few - experimental effects fail to achieve stat significance
  - Too many - statistical significance for effects of no practical value
- Within-subjects aka repeated measures, each test condition are repeated for each participant
  - Advantages: fewer participants, less "variation", no need to balance groups (because only 1)
- Between-subjects separate group of participants for each test condition
  - Advantages: no order effects - disadvantages: more participants, more variation, need to balance groups
- Order effects is offset by counterbalancing: participants divided into groups, test conditions administered in diff order to each group, order of test conditions uses Latin square
- Balanced latin square, each condition precedes and follows each other condition an equal # of times



• Ratio	↑ sophisticated ↓ crude
• Interval	
• Ordinal	
• Nominal	



## CHAPTER 6: HYPOTHESIS TESTING

- Null hypothesis assumes there will be no difference, stats procedures either reject or accept it
- Statistical procedures for hypothesis testing come in two flavors: parametric and non-parametric
  - Parametric: data assumed to come from a distribution, ie normal, t-distribution, etc.
    - ANOVA (analysis of variance), used for ratio+interval data, most common in HCI

Independent variables	Effects					Total
	Main	2-way	3-way	4-way	5-way	
1	1	-	-	-	-	1
2	2	1	-	-	-	3
3	3	3	1	-	-	7
4	4	6	3	1	-	14

- Goal is to determine if IV has sig effect on DV
- Non-parametric: data not assumed to come from distribution
  - Chi-square, used for nominal (ones below are used for ordinal data)
  - Mann-Whitney U, Wilcoxon Signed-Rank, Kruskal-Wallis, + Friedman tests
- ANOVA results: F-statistic, DOF for f-stat, P value
- ANOVA assumptions
  1. population from sample drawn is normally distributed
  2. Independence of cases (sample cases are independent of each other)
  3. Homogeneity of variance: variance among groups is approx equal
  4. ONLY FOR REPEATED MEASURES, sphericity assumption
    - Condition where variances of differences between all possible pairs within-subject conditions (ie levels of IV) are equal
- Chi-square tests investigate relationships between nominal data
  - Data is organized in a contingency table
  - Compares observed values against expected values (these assume no dif)
  - $X^2$  is significant if it exceeds the critical value
- Between-subjects generate independent samples because different Participants are tested with each condition
- Within subjects designs generate correlated samples because the same Participants are tested with each condition

## 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?*

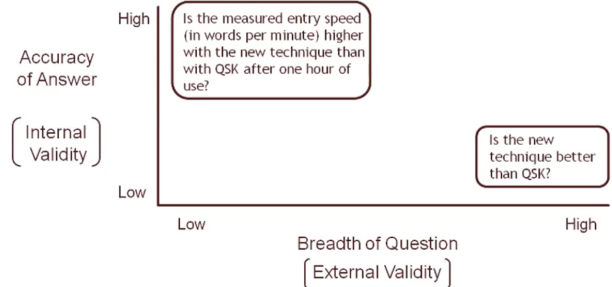
### Factors vs Levels

ANOVA Table for Dependent Variable (units)

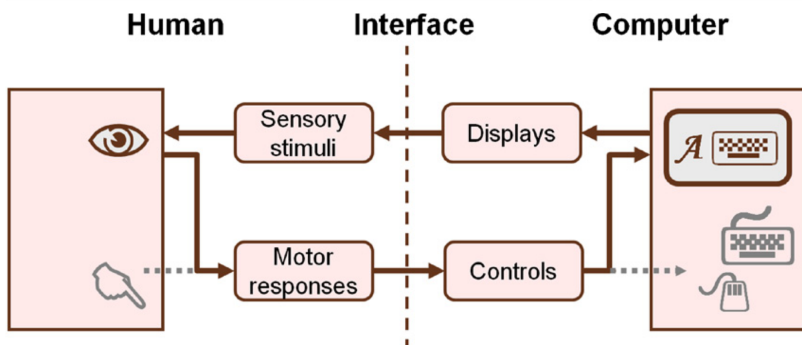
	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	15	81.109	5.407				
Test Condition	3	182.172	60.724	4.954	.0047	14.862	.896
Test Condition * Subject	45	551.578	12.257				

- There was a significant effect of Test Condition on the dependent variable ( $F_{3,45} = 4.95, p < .005$ )
- Degrees of freedom  $p < 0.5$  means significant
  - If  $n$  is the number of test conditions and  $m$  is the number of participants, the degrees of freedom are...
  - Effect  $\rightarrow (n - 1)$  | For this example,  $4 - 1 = 3$
  - Residual  $\rightarrow (n - 1)(m - 1)$  | For this example,  $(4 - 1) * (16 - 1) = 45$
  - Note: single-factor, within-subjects design

## A Tradeoff



Design	Conditions	
	2	3 or more
Between-subjects (independent samples)	Mann-Whitney U	Kruskal-Wallis
Within-subjects (correlated samples)	Wilcoxon Signed-Rank	Friedman



Scale (sec)	Time Units	System	World (theory)
$10^7$	Months		<b>SOCIAL BAND</b>
$10^6$	Weeks		
$10^5$	Days		
$10^4$	Hours	Task	<b>RATIONAL BAND</b>
$10^3$	10 min	Task	
$10^2$	Minutes	Task	
$10^1$	10 sec	Unit task	<b>COGNITIVE BAND</b>
$10^0$	1 sec	Operations	
$10^{-1}$	100 ms	Deliberate act	
$10^{-2}$	10 ms	Neural circuit	<b>BIOLOGICAL BAND</b>
$10^{-3}$	1 ms	Neuron	
$10^{-4}$	100 $\mu$ s	Organelle	



(b)

Operation	Typical time (ms)
Sensory reception	1 – 38
Neural transmission to brain	2 – 100
Cognitive processing	70 – 300
Neural transmission to muscle	10 – 20
Muscle latency and activation	30 – 70
Total:	113 – 528

Measurement Scale	Defining Relations	Examples of Appropriate Statistics	Appropriate Statistical Tests
Nominal	• Equivalence	• Mode • Frequency	• Non-parametric tests
Ordinal	• Equivalence • Order	• Median • Percentile	
Interval	• Equivalence • Order • Ratio of intervals	• Mean • Standard deviation	• Parametric tests • Non-parametric tests
Ratio	• Equivalence • Order • Ratio of intervals • Ratio of values	• Geometric mean • Coefficient of variation	