Interactive Systems (H) 2021

Alistair Morrison
alistair.morrison@glasgow.ac.uk

Timetabled classes

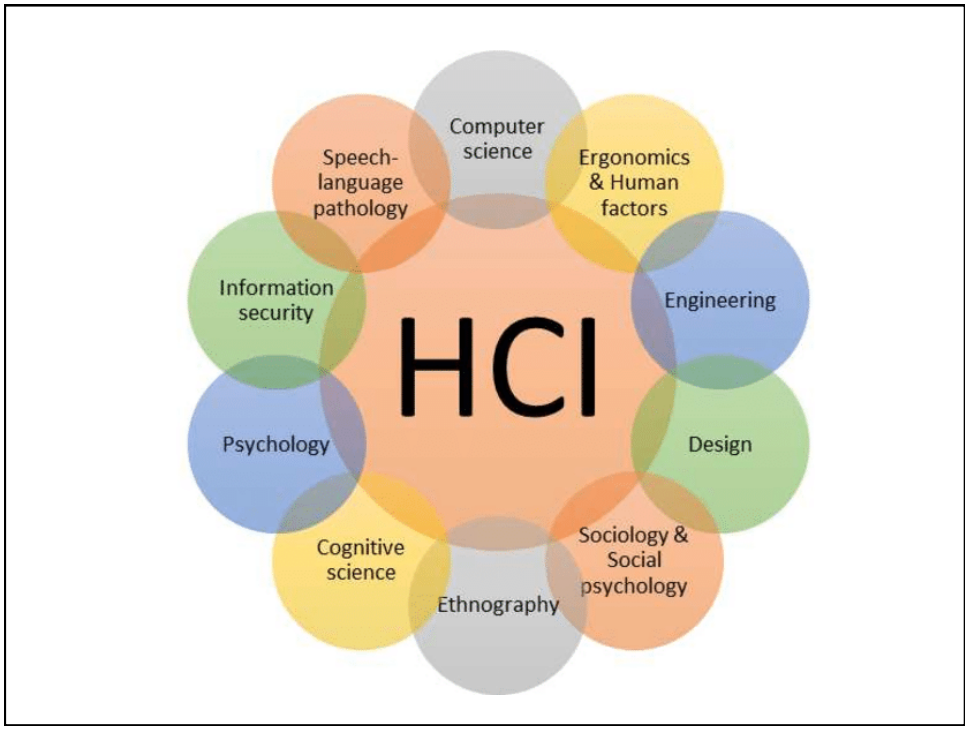
- Weekly lectures on Moodle at start of each week
  - Available for duration of course
- Live sessions
  - 10am – 12pm every Friday
  - Live sessions will have quizzes, polls, discussions
    - But **nothing assessed** during live lectures
- Weekly 1-hour labs
  - Slots 1pm, 2pm, 3pm or 4pm every Friday
  - Exercise sheets will be available at start of week
    - Try to work on them before the lab; prepare questions

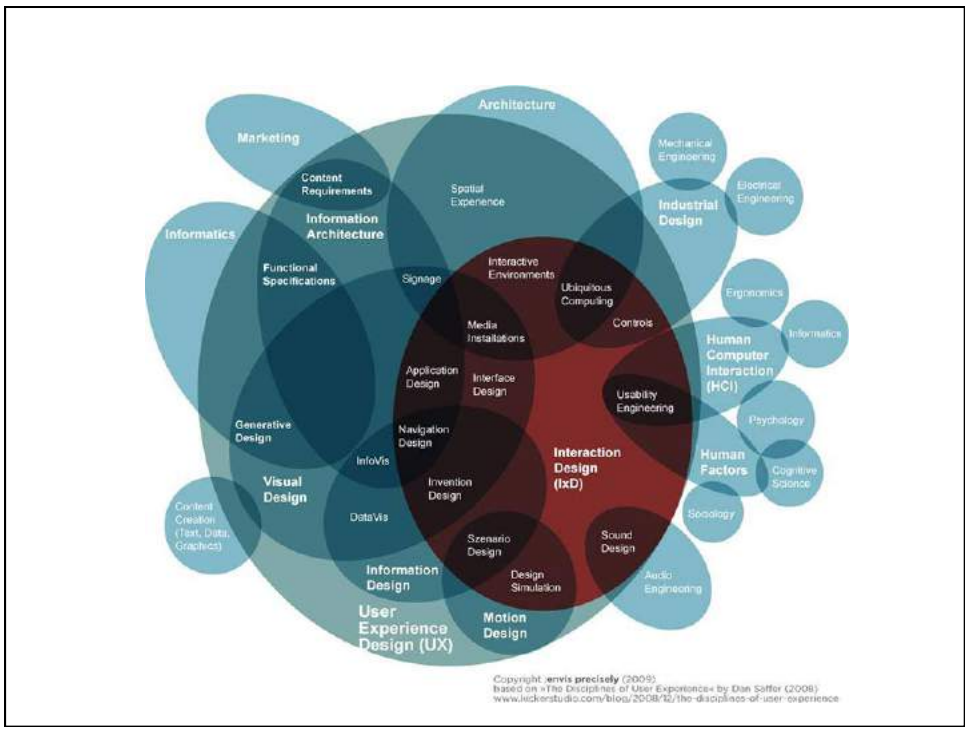
Course materials

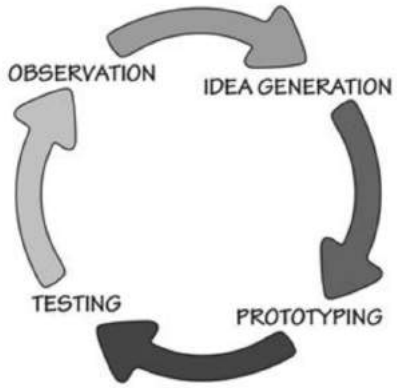
- Moodle section for each week
  - Required reading
  - Lecture notes (these slides)
  - Lectures divided into 20-30 min videos
  - Link to Zoom for live session
  - Lab sheets
  - Submission point for assessed exercises
- Labs mix of live/online on Microsoft Teams
  - Switches weekly

Interactive Systems

- This course requires a different kind of effort than many other computing courses
- The readings are essential to succeed in this course
- You must be able to articulate and defend your ideas
  - Concepts and terms are crucial to effectively communicate


HCI
Computer science
Ergonomics & Human factors
Engineering
Design
Sociology & Social psychology
Ethnography
Cognitive science
Psychology
Information security
Speech-language pathology


Architecture
Marketing
Mechanical Engineering
Electrical Engineering
Content Requirements
Spatial Experience
Industrial Design
Information Architecture
Informatics
Functional Specifications
Signage
Interactive Environments
Ubiquitous Computing
Ergonomics
Controls
Media Installations
Information
Human Computer Interaction (HCI)
Application Design
Interface Design
Usability Engineering
Psychology
Generative Design
Navigation Design
Interaction Design (IxD)
Human Factors
Cognitive Science
InfoVis
Visual Design
Invention Design
Content Creation (Text, Data, Graphics)
DataVis
Sociology
Scenario Design
Sound Design
Information Design
Design Simulation
Audio Engineering
User Experience Design (UX)
Motion Design
Copyright :envis precisely (2009)
based on «The Disciplines of User Experience» by Dan Saffer (2008)
www.kickerstudio.com/blog/2008/12/the-disciplines-of-user-experience



The Iterative Cycle of Human-Centered Design

Human-Centered Design Cycle

- Observation
  - Understand problems
- Idea generation
  - Draw on knowledge, conventions,
  - But be creative, question everything, break conventions if compelling reasons
- Prototyping
  - Paper/low-tech is often quickest way, and less risk of you becoming 'attached' to specific ideas
- Evaluation
  - Many possible methods, as we will see in this course
- …and repeat

How much time do you spend typing each day?

Typing – general stats & info

- People spend an average of 3.2 hours a day typing
  - Average typing speed of 51.56 words per minute
  - "Fast" typists have an average typing speed of 89.56 words per minute
  - "Fast" typists make fewer errors
  - "Fast" typists use both hands more effectively

How would you design a product to improve typing speed?

Designing Interactive Systems

- What is your idea?
- Why do you think it will work?
- Where is the proof?

- Collecting data from potential users
  - Can provide insights, but must be collected in a valid and theoretically sound way

Observations on Typing from 136 Million Keystrokes

- 168,000 participants transcribed 15 sentences in a timed typing task
  - Each set of 15 sentences was selected randomly from a corpus of 1,525 (see citation below for detail)
- Tasked to type as quickly and accurately as possible
- Each keystroke recorded and precisely timestamped
  - including delete keys

Now what?

https://doi.org/10.1145/3173574.3174220

Observations on Typing from 136 Million Keystrokes

- Metrics For Evaluating Typing
  - Performance Measures
    - Words Per Minute
    - Uncorrected Error Rate
    - Error Corrections
    - Keystrokes Per Character
    - Inter-key Interval
    - Keypress Duration
  - Error Metrics
    - Substitution Error Rate
    - Omission Error Rate
    - Insertion Error Rate

Observations on Typing from 136 Million Keystrokes

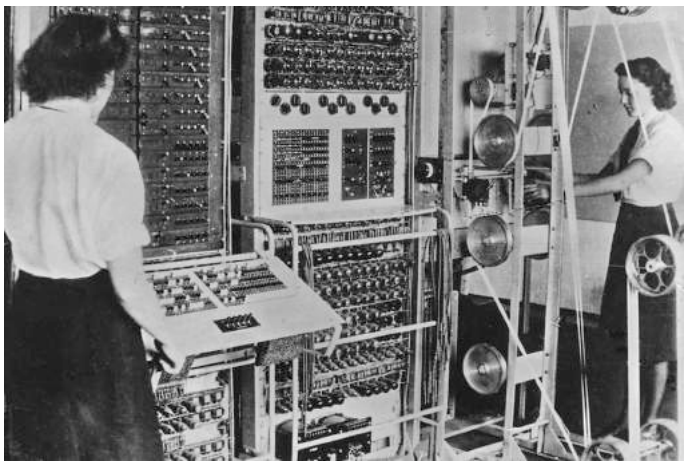
- The use of rollover typing, where the previous key is not released before the following key is entered
- On average, users perform rollover for 25% of keystrokes (SD = 17%) and this has a high correlation with performance ($r = .73$, $p < 0.001$)
  - Fast typists perform rollover for 40-70% of keystrokes
- So, an idea:
  - If we could encourage more rollover on soft keyboards, we might see faster typing speeds
  - Design a soft keyboard that employs haptic feedback to train and encourage rollover typing behaviours

By the end of this course, you will be able to solve this problem

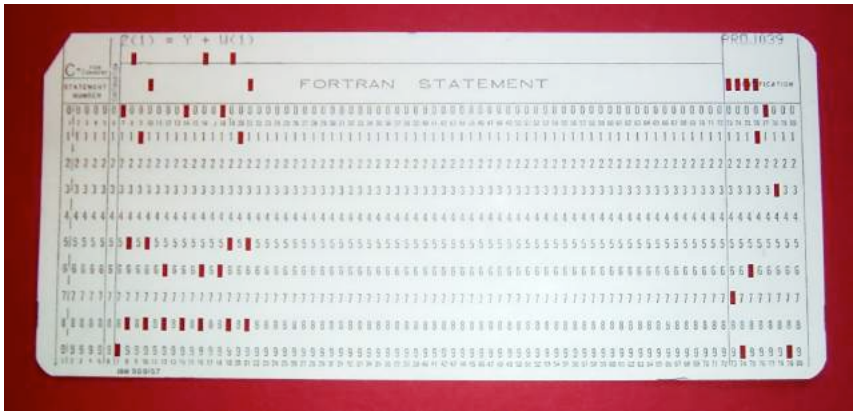
And hopefully others like it.

History of HCI

- Tied to history of Computing
- Initial computers in research labs, took up full rooms
  - Only ever operated on by specialists/engineers/the people who built them
- As technology progressed, got smaller/more affordable, started appearing in workplaces and homes
  - A need for 'real' people to be able to operate them
  - Thoughts of human efficiency/task completion times/ error rates
- A need for a new discipline to study these issues



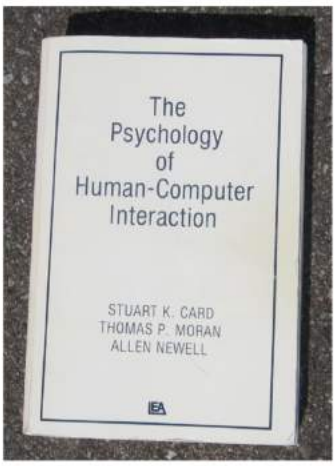
Colossus (1940s) − Bletchley Park code breaking



Programming via punched cards

History of HCI

- HCI generally thought of as beginning in early 1980s
  - (Although studies done before, that in retrospect followed 'HCI' methods and principles)
  - Conferences began
  - Influential textbooks
  - Emergence of the Graphical User Interface



- Trying to create an applied psychology of HCI
- Based on knowledge from human psychology
  - Perception
  - Cognition
  - Motor function
- For many in software design communities, first exposure to Psychology basics
- Engineering-style theories to give approx calculations of how efficiently humans would interact

Xerox Star

Graphical User Interfaces

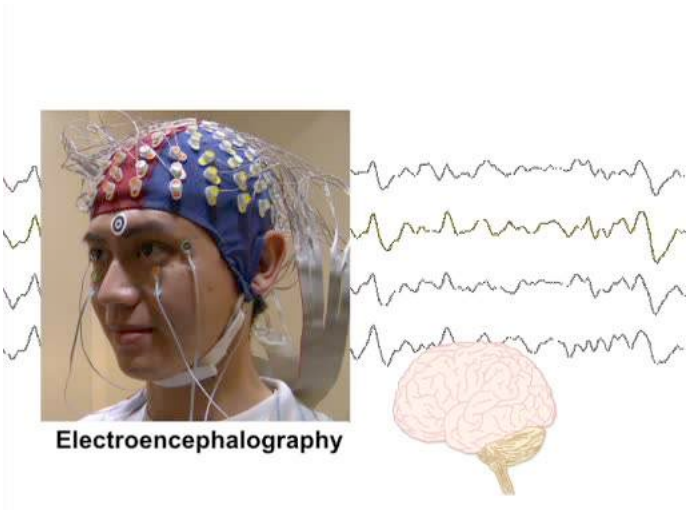
- Xerox Star − 1981
  - The first GUI computer released
  - Bit-mapped display
  - WIMP, WYSIWYG
  - Desktop metaphor
  - Yet not a commercial success
    - Very expensive; network terminal, not 'personal' computer
- Apple Macintosh − 1984
  - Brought the GUI to a wider audience

Broadening of HCI topics

- ~1980s: early research often looking at efficiency
  - e.g measure speed and accuracy
  - Lab-based studies
  - Formal experiments
- ~1990s: field started to broaden, alongside importance of Internet
  - Emails, Web: topics related to communication
- ~2000s: mobile/portable computing
  - Real world studies 'in the wild'
  - New technologies: sensors, wearable, VR/AR…
  - Study social, emotional, cultural issues
  - "Older" forms of research haven't gone away!

Broadening methods

- Technology pushed progress here too
  - Eye-tracking studies, EEG...
  - Large-scale studies, users' own devices
- From early studies that timed tasks/counted errors
- Brought in techniques more from sociology than psychology
  - Ethnography
  - Interviews
  - Case studies



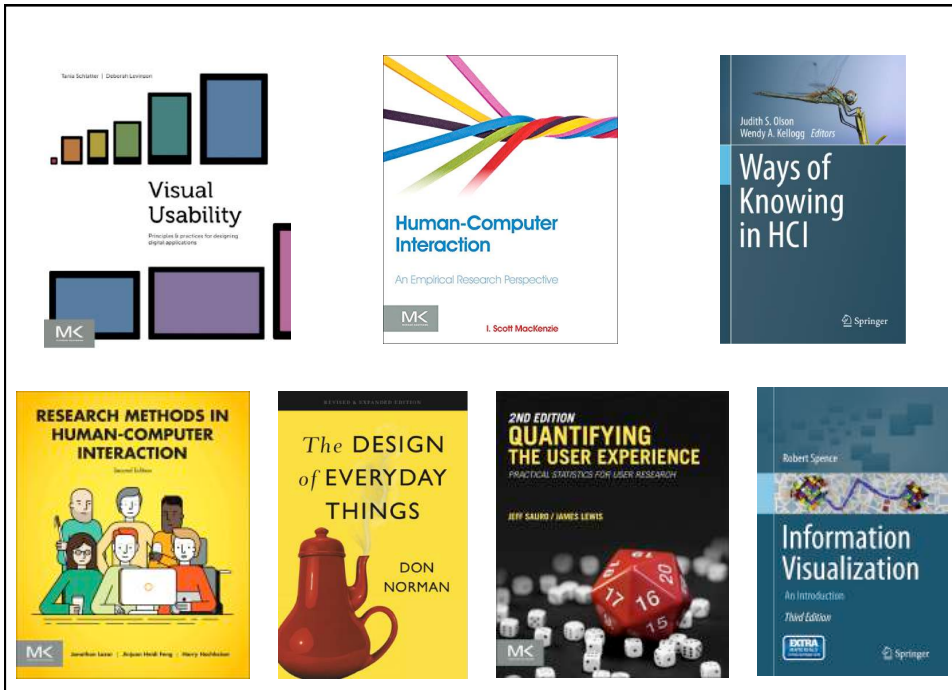
HCI has progressed in three “Waves”

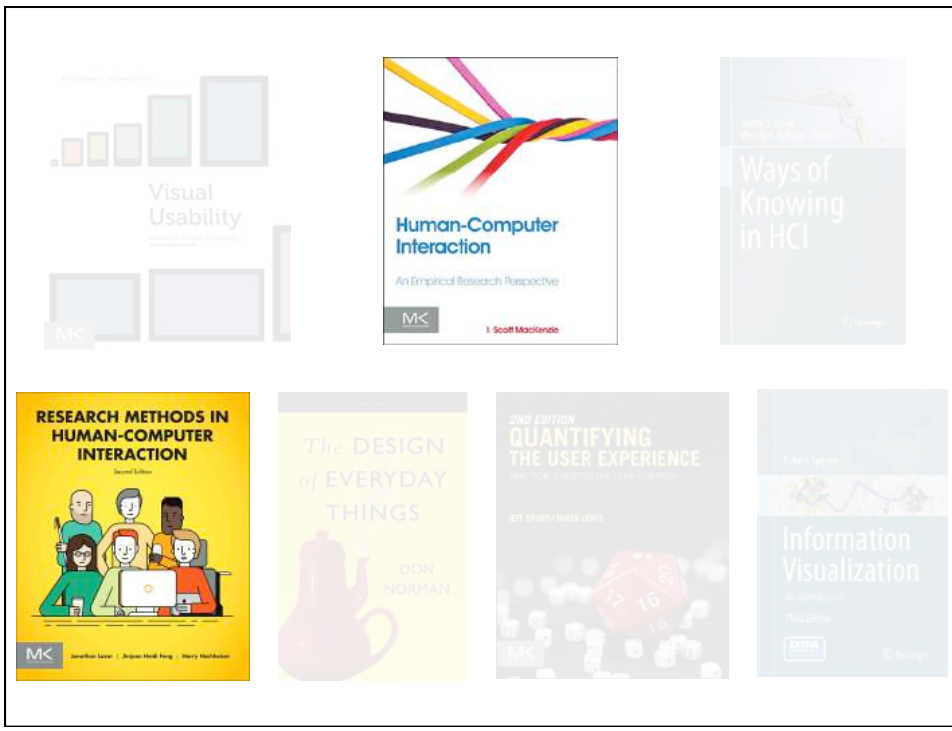
- First Wave: Psychology and Perception
- Second Wave: Organisational and Process Oriented
- Third Wave: Social and Ubiquitous
- Fourth Wave?

For Reference: https://dl.acm.org/citation.cfm?id=1182476

IS(H) Topics

- Visual Design
- Human Perception and Capabilities
- Experimental Design
- Surveys, Focus Groups, and Ethnography
- Analysis Techniques and Statistics
- Modelling interactions
- Large-scale trials & analysis
- Information visualisation


Visual Usability
Human-Computer Interaction
An Empirical Research Perspective
I. Scott MacKenzie
Judith S. Olson
Wendy A. Kellogg Editors
Ways of Knowing in HCI
Springer
RESEARCH METHODS IN HUMAN-COMPUTER INTERACTION
The DESIGN of EVERYDAY THINGS
DON NORMAN
2ND EDITION
QUANTIFYING THE USER EXPERIENCE
PRACTICAL STATISTICS FOR USER RESEARCH
JEFF SAURO / JAMES LEWIS
Robert Spence
Information Visualization
An Introduction
Third Edition
EXTRA


Visual Usability
Human-Computer Interaction
An Empirical Research Perspective
I. Scott MacKenzie
Ways of Knowing in HCI
RESEARCH METHODS IN HUMAN-COMPUTER INTERACTION
The DESIGN of EVERYDAY THINGS
DON NORMAN
2ND EDITION
QUANTIFYING THE USER EXPERIENCE
Information Visualization

What is Examinable?

- All assigned readings as listed in Moodle
- All text of lecture slides

What is not examinable?

- Links provided in lecture slides labelled as “For Reference”

Assessments

	Deadlines
• (1) Prototype Design & Evaluation (20%)	
– Usability evaluation (5%)	15 Oct
– Design & prototype (5%)	29 Oct
– Evaluation plan (5%)	5 Nov
– Evaluation report (5%)	12 Nov
• (2) HCI experiment data analysis (20%)	
– Stats & visualisations from real-world data set	3 Dec
• (3) Exam (60%)	Dec (online)

AE 1: Prototype Design and Evaluation

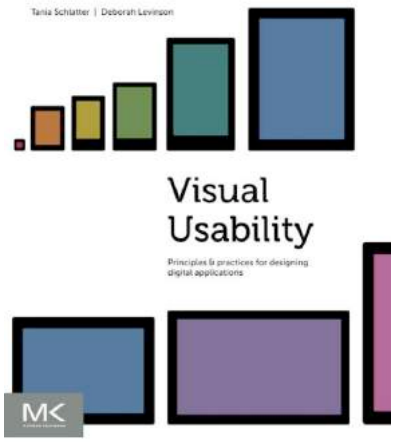
- Split into 4 sub-tasks, with 4 deadlines
  - Usability assessment, prototype, evaluation plan, evaluation
  - There will be a lab sheet on Moodle for each stage with more detailed explanation
- Practical exercise, work on in labs
  - Marks for each stage are awarded out of 4
  - Feedback can be requested from tutors during labs
- Work in lab groups; submit each stage via Moodle
- Submit team members' contribution %s at end of AE1

AE 1-1: First stage

- Conduct a usability evaluation
- AE1-1 Exercise sheet will be released at end of week 2 (so you don't get confused with this week's lab)
- Work on exercise before/during lab next week – **8 Oct**
- Deadline to submit PDF of work in Moodle following week – **15 Oct**
- Practical exercise that will be marked out of 4
- (this will be general pattern of all AE1 stages)

AE 2: Analysis Project

- Reproducing an HCI analysis based on an openly available dataset of typing
- Use the Python notebook template provided, complete all sections
- **Individual** exercise
  - No team component to AE2!
- Submit a PDF of your completed notebook (do not submit an ipynb file)
  - Deadline on Friday in week 11



Visual Usability

by Tania Schlatter & Deborah Levinson

Visual communication of any kind, whether persuasive or informative ... should be seen as the embodiment of form and function: the integration of the beautiful and the useful."
—Paul Rand, A Designer's Art, p. 3

The Things You See Around You Today Are Not There by Random Chance

The interfaces familiar with us may seem easy to design, but are the result of many attempts and many failed designs.

Visual Usability

- Multidisciplinary challenges
  - Graphic design alone doesn't help teach us how to create complex interactive systems
  - Usability alone might not teach us how to create the best experiences
- A complex interface might need to convey many messages
  - Should provide order, patterns to help people process information and derive meaning
- Visual usability
  - Designs grounded in principles of aesthetics and understanding of people
  - We should be able to design and defend a design based on heuristics and best practices

Consistency

- Establishing consistency means setting and maintaining expectations
- External Consistency
  - Consistency with other similar applications
  - If you are designing an interface for online shopping, it should be similar to the established look/feel of existing interfaces
- Internal Consistency
  - Consistency within different parts of an application
  - If you are designing an interface for online banking, all the views need consistency
- Internal/external can sometimes clash
  - E.g. suite of apps from same company. Should they primarily look and feel like each other, or should each one meet the conventions of that type of app?

Types of Consistency

- Layout Consistency
  - Screens showing similar information should have all elements positioned the same way
  - Spatial relationships should remain consistent
- Typography Consistency
  - Use fonts, weights, and sizes meaningfully and consistently
- Colour Consistency
  - Use colour to establish and maintain consistency - typically means establishing a defined colour scheme
- Imagery Consistency
  - Charts, logos, videos, photography, icons, backgrounds, and anything else that isn't typography

Breaking consistency

- Can break the rules to make a point/highlight something
- e.g. make one piece of content bigger than others if it's the most important/where you want to guide the eye
- Don't change more than two aspects of a single item

Hierarchy

- Visual hierarchy is used to communicate structure, relationships, and relative importance
- More important items need more “visual weight”
  - Understanding behaviour of gaze is important when deciding how to effectively give important elements more visual weight
  - Use position, size, colour, groupings, contrast, control types to represent priorities
  - Make sure people notice what they need to − based on identified user priorities
- Start with black and white wireframes − only vary size and positions

Layout

- Screen size
  - The screen gives the frame within which your entire interface sits
  - Core layout principles might apply to all screen sizes − main thing is how elements relate to each other, so layout can flex
- Position
  - Does the relative position of elements communicate structure
  - Might need to balance lots of relationships
- White Space
  - Absence of content (not necessarily ‘white’!) is equally important, for example the sparse design on a Google landing page
  - Trick to create dense but appealing screens is white space to group and establish hierarchy
- Grid
  - Align items relative to (invisible) horizontal and vertical lines

Proximity, Scale, and Alignment

- Proximity
  - Is the relative placement of items arbitrary or meaningful?
- Scale
  - Is the relative size of elements arbitrary or meaningful?
- Alignment
  - Is the alignment consistent and used to represent the hierarchy?

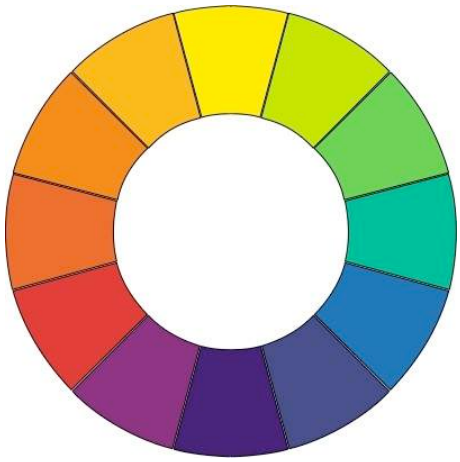
Colour

- Powerful way of attracting the eye
  - Can create emotional response
  - Enhance usability and appeal
  - Aid understanding by creating connections between related items
- Colour choices
  - Can be cultural associations
  - Specific UI conventions
    - E.g. Red for error messages. But maybe only if critical
- Shouldn't convey anything crucial through colour alone
  - Visually impaired/colour-blind users

Colour properties

- Hue is a categorical description of the perceived colour
  - Red, yellow, green, cyan, blue, violet, magenta, purple
- Saturation is the purity of colour compared to grey
  - When fully saturated, the ‘purest’ form of the hue
  - Saturated colours can draw the eye more
- Brightness - relative amount of light

Colour circle

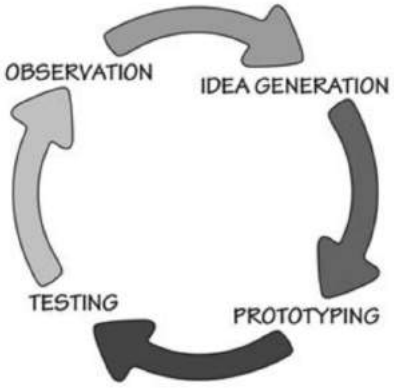


Types of Colour Contrast

- Warm-Cool Contrast

- Complementary Contrast

- Test all design decisions with prototyping
- Iterate!

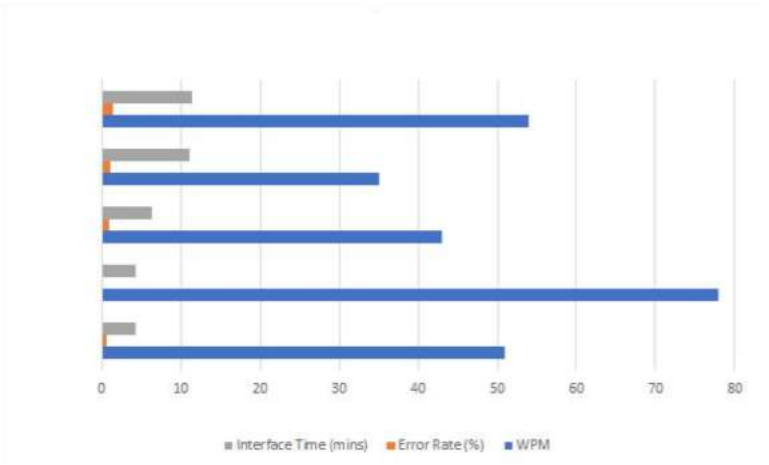


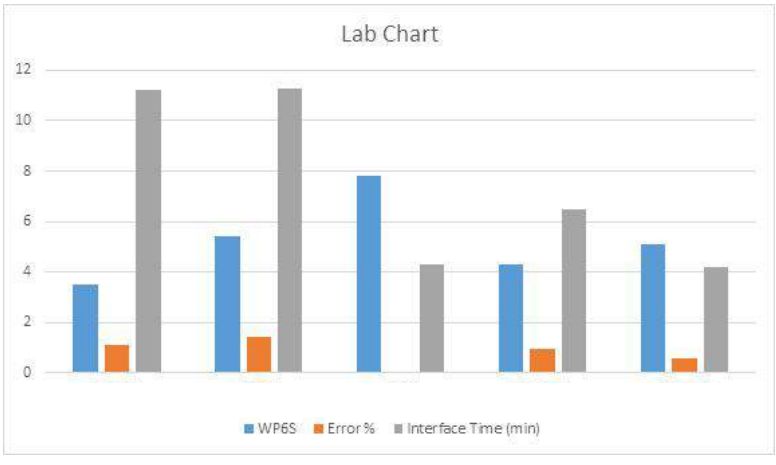
Human Perception and Capabilities

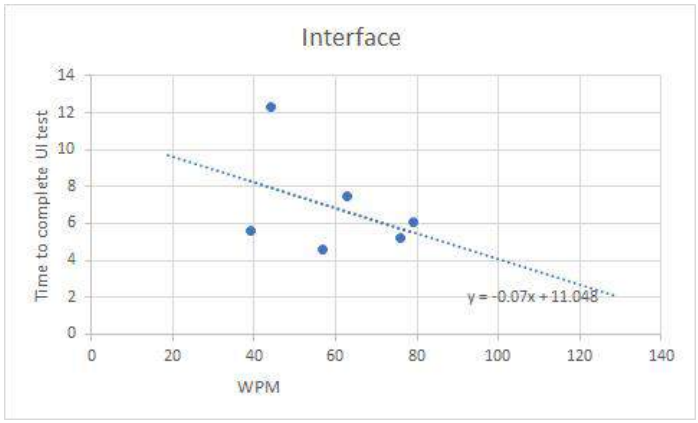
Lecture 3-4

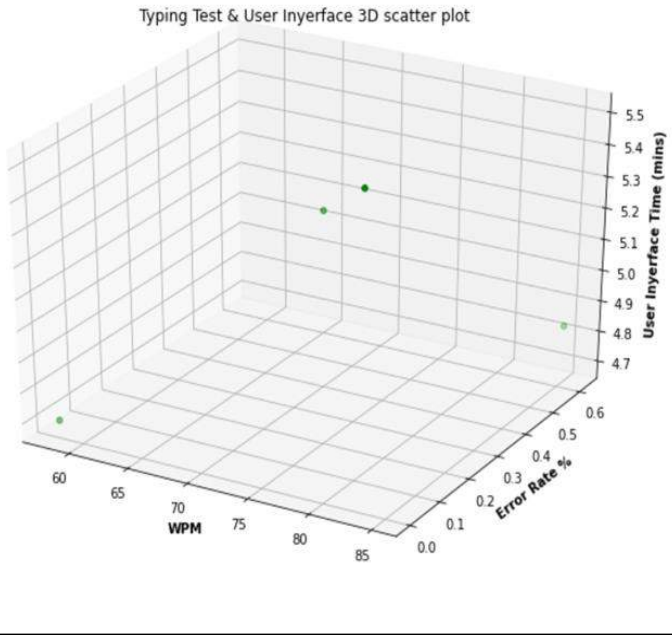
Lab 1 comments

- My observations: most people used Excel
  - Jupyter/matplotlib/seaborn encouraged options
- Many possibilities for what to plot


0
10
20
30
40
50
60
70
80
Interface Time (mins)
Error Rate (%)
WPM

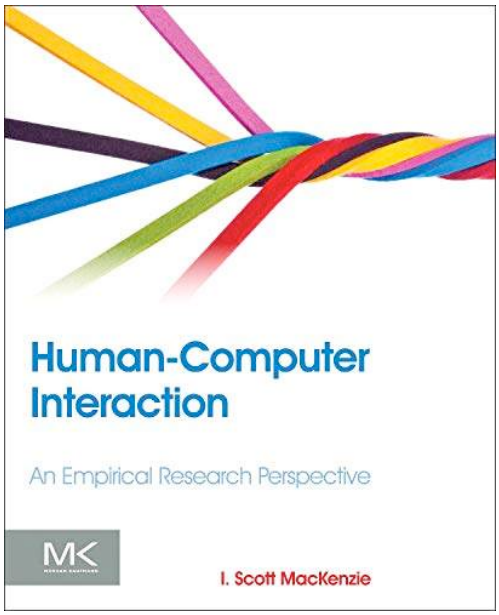

Lab Chart
12
10
8
6
4
2
0
WP6S
Error %
Interface Time (min)


Interface
Time to complete UI test
WPM
y = -0.07x + 11.048
0
2
4
6
8
10
12
14
0
20
40
60
80
100
120
140


Typing Test & User Inyerface 3D scatter plot
WPM
Error Rate %
User Inyerface Time (mins)
60
65
70
75
80
85
0.0
0.1
0.2
0.3
0.4
0.5
0.6
4.7
4.8
4.9
5.0
5.1
5.2
5.3
5.4
5.5

Studying the Human

- HCI: **Human** Computer Interaction
- Early HCI work took findings/approaches from Psychology to apply to interactions with computers
  - Perception
  - Cognition
  - Motor function
- Used to guide system development
- Continue to measure, refine, experiment



Time Scale of Human Action

- Social Band
- Rational Band
- Cognitive Band
- Biological Band
  - Less relevant for most HCI research/practice you will be part of

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

Cognitive Band

- 100 milliseconds to 10 seconds
  - Pointing devices, selection techniques, text entry, gestural input
  - Times based on reaction times and biomechanical properties
- Consider how users perform multitouch rotation gestures
  - Does the angle of rotation impact performance?
  - Do users pivot from the thumb or rotate multiple touchpoints?
  - Does the starting angle impact performance?

For reference: https://dl.acm.org/citation.cfm?id=2481423

Rational Band

- Occupy minutes or hours
  - Tasks, like web site use, user search strategies, OS navigation
  - Users must experience an interface and **make decisions** about their next actions
- Consider an evaluation of user search behavior
  - How often do users "branch" their search results?
  - How many "branches" do users generate during a typical search?
  - Why do users establish a new "branch"

For reference: https://dl.acm.org/citation.cfm?id=2124322

Social Band

- Days, weeks, months
  - Activities such as as workplace habits, social networking, online dating, privacy
  - Require development of social bonds or establishing norms/ standards
- Consider a study on how people develop relationships in online dating
  - Interviews with members of the community
  - Participation/observation in active forums
- Qualitative methods dominate
  - Although often opportunity for mixed methods studies/data analytics too
    - For Reference: https://theblog.okcupid.com

For reference: https://dl.acm.org/citation.cfm?id=2702417

Model of Human Computer Interaction

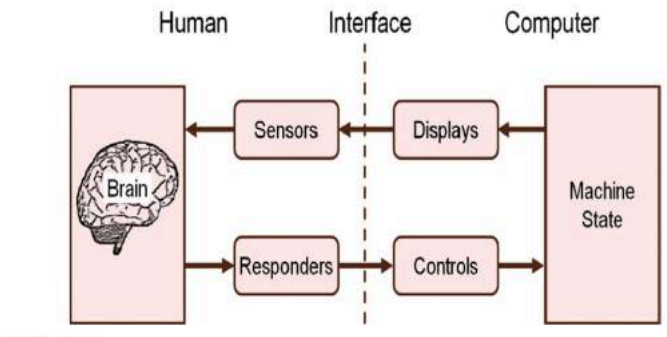


FIGURE 2.2

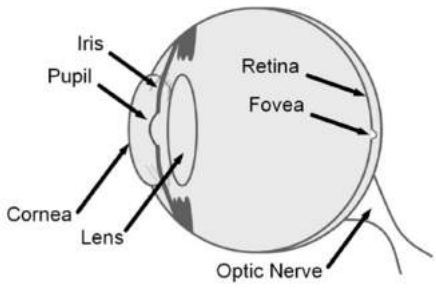
Human factors view of the human operator in a work environment.

(After Kantowitz and Sorkin, 1983, p. 4)

Human Senses

Vision

- Light passes through the lens
- The lens focuses light into an image projected onto the retina
- The retina converts visible light into neurological signals
- The centre of the retina, the fovea, processes details



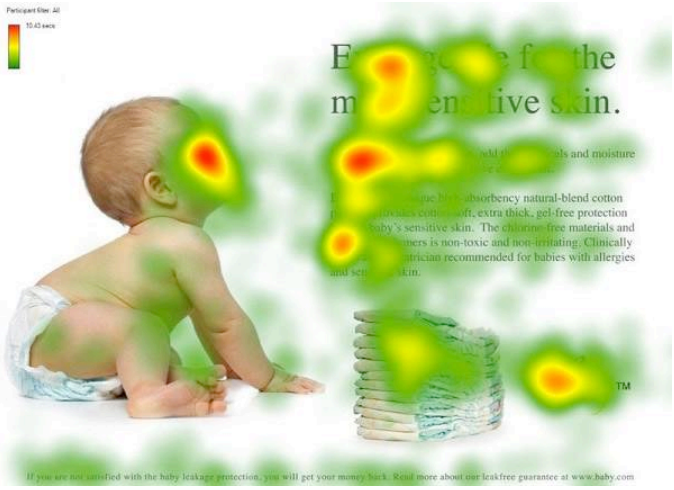
Properties of Vision

- Frequency of visible light
- Intensity
  - Eye light sensitivity varies by wavelength
- Fixations and Saccades
  - Fixations processes detail while the eyes are still
  - Saccades are rapids movements (30-120 ms) of the eyes to a new position
- Understanding human eye motion is important to understand how content is viewed
- Scanpath
  - In eye tracking studies, a sequence of fixations & saccades

Eye-Tracking

Eye-Tracking

Eye-Tracking



Hearing

- Sounds are perceived from cyclic fluctuation of pressure
  - Typically in air
- Loudness
  - Subjective perception of sound pressure level
- Pitch
  - Subjective perception of frequency
- Timbre
  - Harmonic structure to be described as richness or brightness
- Envelope
  - Changes in the subjective properties over time

Touch

- Touch / haptic
  - Through vibration, air, and ultrasound
  - For reference: https://dl.acm.org/citation.cfm?id=2663280
- Temperature
  - For reference: https://dl.acm.org/citation.cfm?id=1979316
- Pain
  - We try to avoid this in HCI
- Proprioception
  - The ability to sense the position of your body and limbs

Smell

- Olfaction
  - The ability to perceive odours
- HCI has explored scent through scent "cubes"
  - fans that disperse scent, and pressurized delivery systems
- Olfoto: tagging photos with smells vs text
  - For reference: https://dl.acm.org/doi/abs/10.1145/1124772.1124869

Taste

- Yes, really!
  - Chemical reception of sweet, salty, umami, bitter, and sour
- TastyFloats
  - Levitate food onto user's tongue
  - For reference: https://dl.acm.org/citation.cfm?id=3134123
- Summary of multi-sense interactions
  - For reference: https://dl.acm.org/citation.cfm?id=3134123

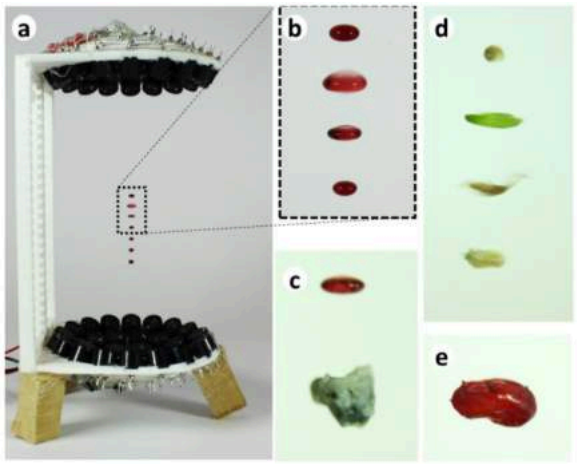


Figure 1. Examples of levitated food morsels: a, b) Acoustic levitation of droplets of wine; c) Wine and blue cheese; d) Bread, lettuce, meat and bread; e) and a raspberry grain.

Human Responses

Limbs

- Input for systems is primarily achieved by moving the limbs in 3D space
  - Think of typing, using a mouse, using a trackpad
  - We use our limbs to generate a signal that is interpreted as input

Voice

- Speech recognition has come a long way, but we still face challenges of segmentation, recovery from errors, and information throughput

Eyes

- Selection based in Gaze is a common approach in VR, and becoming more common in less instrumented environments as well
  - For example, consider common phone unlocking techniques
- Most info probably also coming in through vision, so eyes doing double tasks

Human Brain

- Connects the sensors and responders
- Different processes/capabilities
  - Perception
  - Cognition
  - Memory

Perception

- First stage of processing in the brain
  - Associations and meanings take shape
- Just Noticeable Difference
  - Below what threshold can humans no longer perceive difference?

For reference: https://dlnext.acm.org/doi/abs/10.1145/2556288.2557033

Perception and Ambiguity

- Illusions work when our perception fills the gaps in ambiguous stimuli
  - Ponzo lines demonstrate how our depth perception changes how we look at the two black lines
- There are illusions that can trick our visual, aural, and haptic senses

Cognition

- Human process of conscious intellectual activity
  - Thinking, reasoning, deciding

Memory

- Ability to store, retain, and recall information
- Short term memory capacity 7 ± 2
  - Has often been used to guide UI design, eg number menu items
    - Might be misunderstanding the original intent
    - Shorter menus probably still good though!

Human Performance

Human Performance

- Speed Accuracy Tradeoff – tasks completed faster are more error prone
  - People often prioritise speed or accuracy differently based on context
- Most of early HCI measured this, but still important and studied today
- E.g. performance with various input devices
  - But also augment overall human performance, e.g. find answers to questions with visualisation tool vs looking at raw numbers

Reaction Time

- Different sensory modalities have different reaction times
  - 150ms audio
  - 200ms visual
  - 300ms smell
  - 700ms pain
- Visual search is another example of reaction which includes more complex cognition that simply responding to stimuli

Skilled Behaviour

- In most tasks beyond simple responses, human performance can increase with practice
  - Like playing darts and playing chess
- Playing darts requires training of your sensor/ motor skill
- Playing chess requires training of your mental skill
- Some tasks require both

Attention

- When task performance degrades with performed simultaneously with another, we can say that task requires attention
  - Consider walking and talking
  - Consider reading and talking
  - Driving and talking?

Attention

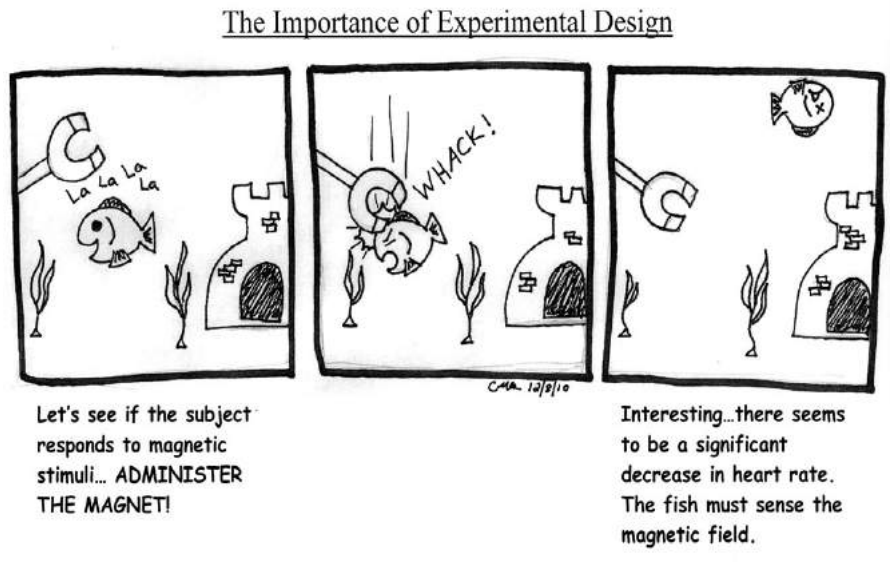
- Divided attention is concentrating/performing more than one task at a time
  - Typically this will degrade performance, which is not an option in safety critical contexts like driving
- Focused attention is attending to one task to the exclusion of all others
  - The ability to ignore external events not always possible or feasible
  - In a noisy room, you might be able to have a conversation but are likely to be distracted at times.
- Sensory modalities are often thought of as channels, but in practice it is not so simple

Human Error

- Error is a discrete event in a task where the outcome is incorrect or deviates from the desired outcome
- In practice, this kind of coarse measure of error provides a limited understanding
  - Consider the Key Stroke dataset
  - “error” in this sense isn’t even reported in this dataset, although it’s simple to calculate
- We are often trying to measure something more complicated than % of errors

Lab Evaluation and Quantitative Methods

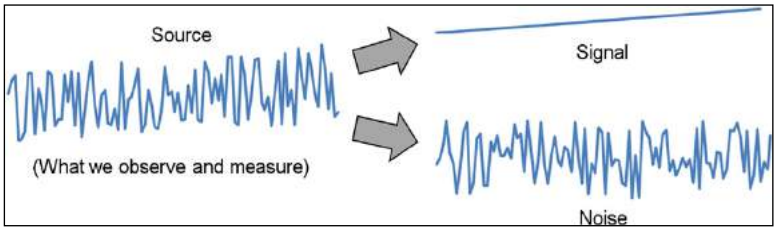
Lecture 5-6



Method / Methodology

Methodology is the way an experiment is designed and carried out.

Sound methodology is critical to allow us to understand what is really going on (signal) in a noisy and messy world (noise):



"Science is method. Everything else is commentary." Allen Newell

Why does this matter?

- Will help you run good studies in your 3rd year projects and 4th year dissertation
- If you want to publish a scientific paper testing hypotheses you need to understand experimental design
- Even if you don't end up running studies (e.g. you go into data science field) you need to know how the data was collected in order to understand how to handle the data
- It's a good life skill – it gives you extra tools to think critically when you read the latest science story such as 'chocolate can cure cancer.' Sound methodology often separates good science from the bad.

Validity

- Internal validity
  - Is effect observed due to varied condition(s)?
- External validity
  - Are experimental results generalisable to other people/situations?
    - Sampling
    - Realistic conditions
- Often tension between interval vs external: one at expense of other?

Ethics

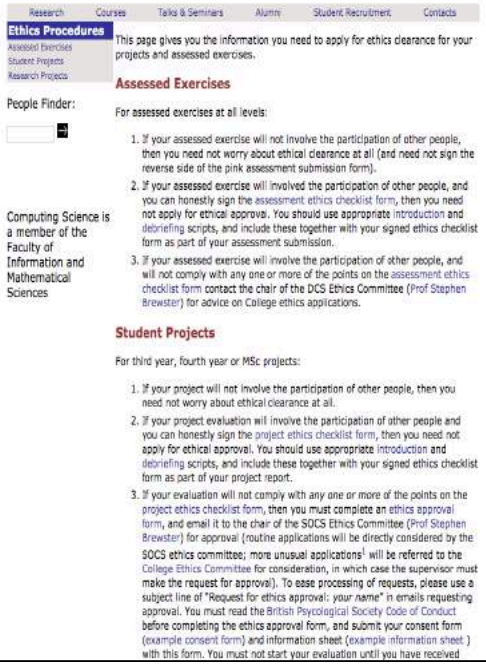
Ethics are important in HCI research.
Often borrow from psychology research

Crucial notion of **informed consent**. Inform participants about:

- Nature of research
- Methodology
- Risks or benefits
- Right not to participate or to withdraw
- Right to anonymity and confidentiality

Particular issues in HCI work can involve recruitment of vulnerable groups (e.g. when investigating assistive technology), or deception that might be involved during a study.

http://www.dcs.gla.ac.uk/ethics/



Research Courses Talks & Seminars Alumni Student Recruitment Contacts

Ethics Procedures

Assessed Exercises
Student Projects
Research Projects

People Finder:

Computing Science is a member of the Faculty of Information and Mathematical Sciences

This page gives you the information you need to apply for ethics clearance for your projects and assessed exercises.

Assessed Exercises

For assessed exercises at all levels:

1. If your assessed exercise will not involve the participation of other people, then you need not worry about ethical clearance at all (and need not sign the reverse side of the pink assessment submission form).
2. If your assessed exercise will involved the participation of other people, and you can honestly sign the assessment ethics checklist form, then you need not apply for ethical approval. You should use appropriate introduction and debriefing scripts, and include these together with your signed ethics checklist form as part of your assessment submission.
3. If your assessed exercise will involve the participation of other people, and will not comply with any one or more of the points on the assessment ethics checklist form contact the chair of the DCS Ethics Committee (Prof Stephen Brewster) for advice on College ethics applications.

Student Projects

For third year, fourth year or MSc projects:

1. If your project will not involve the participation of other people, then you need not worry about ethical clearance at all.
2. If your project evaluation will involve the participation of other people and you can honestly sign the project ethics checklist form, then you need not apply for ethical approval. You should use appropriate introduction and debriefing scripts, and include these together with your signed ethics checklist form as part of your project report.
3. If your evaluation will not comply with any one or more of the points on the project ethics checklist form, then you must complete an ethics approval form, and email it to the chair of the SOCS Ethics Committee (Prof Stephen Brewster) for approval (routine applications will be directly considered by the SOCS ethics committee; more unusual applications[1] will be referred to the College Ethics Committee for consideration, in which case the supervisor must make the request for approval). To ease processing of requests, please use a subject line of "Request for ethics approval: *your name*" in emails requesting approval. You must read the British Psycological Society Code of Conduct before completing the ethics approval form, and submit your consent form (example consent form) and information sheet (example information sheet) with this form. You must not start your evaluation until you have received

Independent Variables

Also called *factors*

*E*xperiments with independent variables are often called *factorial experiments.*

These can be naturally occurring or directly manipulated by the experimenter

Characteristics:

e.g. of computer interface – input device, feedback modality, display size

e.g. of participants – gender, handedness, expertise

OR ***Circumstances****:*

e.g. background noise, room lighting

Levels: each test condition. E.g. 'mouse' and 'trackpad' are *levels* of independent variable 'input device'

Dependent Variables

In HCI the DV is often a measured human behaviour. The measurement *depends* on what the participant does.

Any observable, measurable behaviour can be a DV:
e.g. typing speed, eye movements, 'negative facial expressions', 'read text events', how respond to questionnaire

You can be creative here to get the measurement that tells you most about the impact of the independent variable.

Effect numbers

You can have (and will often want) more than one IV and DV in a study. However it is important to limit these as the more IVs you have the more comparisons there are.

These increase rapidly:

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
5	5	10	6	3	1	25

FIGURE 5.2

The number of effects (main and interaction) increases as the number of independent variables increases.

Increasing IVs quickly accelerate your number of observed effects, so **limit to 1 – 3**.

Control Variables

These are not under investigation (i.e. are not IVs) but they might influence participant behaviour (DVs)

e.g. keyboard angle, chair height, display size

Experimenters control these variables to prevent their influence by setting up their study in a controlled environment and recruiting with strict inclusion / exclusion criteria.

e.g. Right handed only, Experienced users only

Increases internal validity but reduces external

Random Variables

Instead of attempting to control for everything, it is often better to allow some variables to vary randomly in order to generalise the results (and because controlling for everything is probably impossible).

e.g. height, hand/finger size, social disposition

Each study will require judgement about the trade off between control and allowing random variation.

e.g. using a questionnaire of motion sickness and recruiting only those under a threshold:

1) In a study investigating the acceptability of two in-car interaction techniques for the general population

2) In a study comparing the use of a VR headset in a moving vehicle with different VR conditions to mediate sickness level

Confounding Variables

In some cases a circumstance or condition will change ***systematically*** with the independent variable.

e.g. practice, different types of measurement for levels of the IV, prior experience with an interface (e.g. when comparing Google to anything)

Such variables are confounding because they prevent the possibility of a cause and effect relationship being inferred from the results. So they need to be controlled for. **This is a central skill in experimental design**.

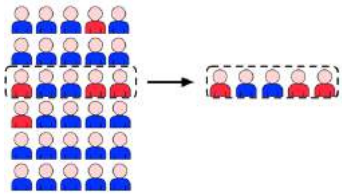
Prior experience will almost definitely confound our attempts to find the best keyboard…

Participants

In order to correctly assume that research results apply to people other than those recruited you must:

Recruit people from the population you want to investigate

Recruit a sufficient number of participants

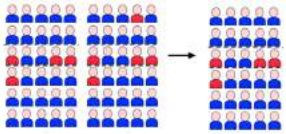


Participants

Recruitment methods

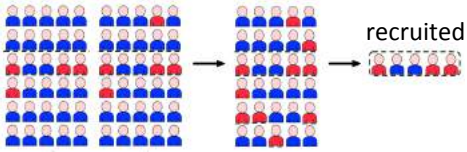
Ideal = participants drawn at random from a population

real pop → recruited sample



In practice = convenience sampling

real pop → convenient pop → recruited sample



How might these people differ from the population you are trying to generalise your findings to?

Participants

How many people should you recruit?

More is better

However there is a balance between representing the population and practical, and sometimes ethical considerations

Practical: Not much time to recruit (e.g. in student project), population difficult to access, more testing delays product going to market

Ethical: Study puts participants under burden; continuing study beyond necessary delays useful intervention or technique that can improve access.

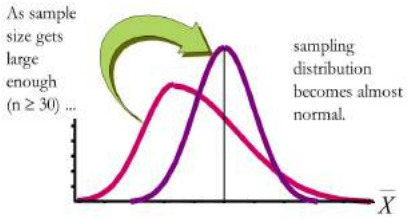
Participants

Central Limit Theorem:

As sample size increases to ≥ 30 becomes approximately normally distributed.

This means that, if you are measuring abilities that are normally distributed in general population (e.g. typing accuracy on a task) then after around 30 people are recruited this will be reflected in the data.

Therefore there has been a rule of thumb in psychology research that 30 people is a good number



Within or Between subjects

Take the example comparing the two keyboard types:

You have 20 willing participants for a study, a typing task and way of calculating typing speed and accuracy. You could approach the comparison in two ways:

Within-subjects (repeated measures)

All 20 participants use both keyboards and you compare the average difference in performance between the two.

Between-subjects

You split the participants into two groups of 10 and compare performance of QWERTY group to Dvorak group

Order Effects

In general within-subject designs are favoured in HCI due to the nature of the research – this means we need to do something about the interference between test conditions or ***order effects:***

Learning effects

Order effects are usually a problem because of learning – people's performance on a task improves as the study goes on and this makes it hard to know if observed differences are due to the IV under investigation (e.g. keyboard type) or this confounding variable (learning leading to performance improvement).

Fatigue effects

These are more rare but it is possible that people get worse at a task the longer it goes on due to tiredness or because they lose interest. This is also a type of order effect.

Counterbalancing

Balance the order in which participants do each level

So counterbalancing is really easy if you have 2 levels of an IV, but what if you have more?

A **Latin square** is an *n* x *n* table that allows conceptualisation (and generating) of counterbalanced conditions

Counterbalancing

Latin Squares

Assign users to groups. Each group gets different order of conditions

Ensure equal number of people in each group

Below we can use Latin squares to ensure that each condition in a text entry experiment is presented first for each group:

	swype	standard
Group 1	1	2
Group 2	2	1
Group 3		
Group 4		

swype	standard	prediction
1	2	3
2	3	1
3	1	2

swype	standard	prediction	voice
1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

Counterbalancing versus Randomisation

- When Latin Squares become needlessly complication or impractical, randomisation is another technique to mitigate order effects
  - Consider a survey with 30 questions
  - Question order effects?
  - Full 30 * 30 Latin Square would get complicated
    - Might need a lot of users, to place equal numbers in each group
  - Could just completely randomise

Key point when counterbalancing:

If you are only interested in removing the order effects then you need to make sure that each possible combination is represented an equal number of times.

Don't need to worry about the size of groups that do each combination of condition orders, as long as all group sizes are same

i.e. when Order effect = a confounding variable

If you want to understand order effects, not just cancel them out, then you need a large enough sample of your participants to do each order combination

Group size has to be large enough to compare condition orders

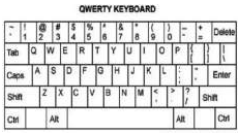
i.e. when Order effect = an Independent variable in the study.

Usually you only want to control for order effects. If that's all you do then you can't make any conclusions about the impact of condition order.

Longitudinal studies

In the keyboard example there is a clear confounding variable – everyone is more likely to be familiar with QWERTY than other keyboard types.

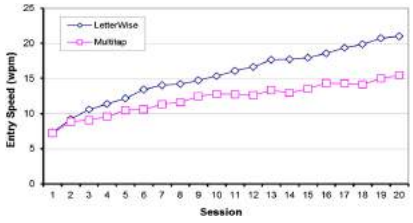
However that is not to say it is the best layout overall, experience level being equal. To test this we need to use a longitudinal study to allow people to become equally experienced with both QWERTY and Dvorak methods.



Longitudinal studies are often used in HCI to investigate learning effects over time. This is very important considering the ubiquity of technology use in everyday life.

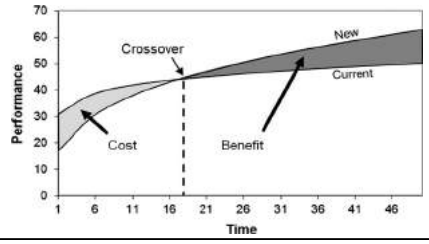
Longitudinal studies

Effects only seen after long term use:



Crossover:

In the case of a new product, it may be that performance on the traditional system will start off better, but that this crosses over after long term use:



Running the Experiment

Always pilot your study before running it!

You won't believe the issues that can arise – technical issues, no-one can understand your task instructions, the study takes 3 hours to finish…

Use consent forms and make sure the participant understands the study

Be consistent

Have a neutral manner and use a script if needed to make sure you give the same instructions each time. Running a research study is an acquired skill.

Also be aware of bias – is your desire for the hypothesis to be met making you inadvertently encourage participants to behave in the desired way?

Surveys, Focus Groups, and Qualitative Methods

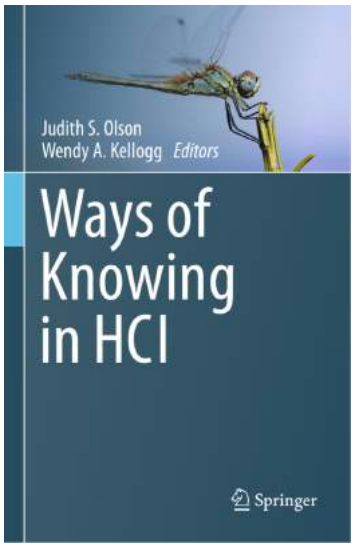
Lecture 7-8

AE1

- More Figma in this week's lab
  - Deadline 29 Oct - but don't have to wait until deadline!
- AE1-3 Plan Evaluation
  - Release 25 Oct
  - Lab 29 Oct
  - Deadline 5 Nov
- AE1-4 Run Evaluation
  - Release 1 Nov
  - Lab 5 Nov
  - Deadline 12 Nov

Surveys

- Surveys allow researchers, designers, and developers to capture high level information about user experiences, attitudes, and perceptions
- Paper / phone / emailed / website
- Low cost, fast, broad reach
- A well designed and analysed survey can provide useful insights; a poorly designed and analysed survey is noise (and a waste of your time and your respondents' time)



Surveys Themselves as a Topic of Research

- Population sampling
- Optimise data collection
  - e.g. return rates
- Reduce biases in questions
- Question order effects
- Computer vs paper-based
  - Early study (1983) found less socially desirable responses & longer open-ended responses in digital survey

Surveys Themselves as a Topic of Research

- Tourangeau's 4 cognitive steps to survey responses (1984):
  - *Comprehension* of the question, instructions, and answer options
  - *Retrieval* of specific memories to aid with answering the question
  - *Judgement* of the retrieved information and its applicability to the question
  - *Mapping* of judgement onto the answer options

What Kinds of Questions are Surveys Good For?

- Attitudes
- Intent
- Task Success
- UX Feedback
- User Characteristics
- Interactions with Technologies
- Awareness
- Comparisons
- Can survey regularly to assess changes over time

What Kinds of Questions are Surveys Bad For?

- Precise Behaviours
  - Log data can often give more accurate info
    - Textbook says “always”. But log data is not infallible
      - If you’re e.g. interested in how often/when people use your app - log data can fail to record, might need to stream back data over unreliable network connection, might record someone else using the device…
- Underlying Motivations
- Usability Evaluations

Survey Pitfalls

- Surveys also need to consider experimental design and confounding factors
  - Common issue is to ask about multiple dependent variables in a single question: On a scale of one to five, rate usability and how enjoyable your experience was
- Low completion rates
  - Repetitive questions, poor usability, poor design
- Noisy data from bad question design
  - Vague or ambiguous questions provide noisy data
- Biased Questions

How to develop a good survey

- Research Goals
- Population and Sampling
- Questionnaire Design
- Review and Testing
- Implementation and Launch
- Data analysis and reporting

Research Goals

- Articulate your research goal and identify the useful *constructs*
  - Example: How acceptable is it to use a VR headset when travelling by air?
    - If we asked this, the data would likely be so noisy it would not tell us anything
    - Constructs: Seat Position, Type of Display
    - Now we ask: How acceptable is it to use a VR headset in the aisle seat? How acceptable is it to use a seatback display in this aisle seat? How acceptable is it to use a VR headset in the window seat?
  - Only crucial constructs: too many 'nice to know' questions and survey might be too long, increase drop-out rate
- Cognitive pretesting: are respondents interpreting constructs as intended?
  - Think-aloud with a few users

Population and Sampling

- As with participants in lab experiments, survey respondents need to be recruited from the target population
  - Refer to lectures 5-6 for details about populations, samples, and convenient samples
- To ensure the intended population is represented, you might develop inclusion criteria
  - For example, only respondents with over 20 hours gameplay can participate

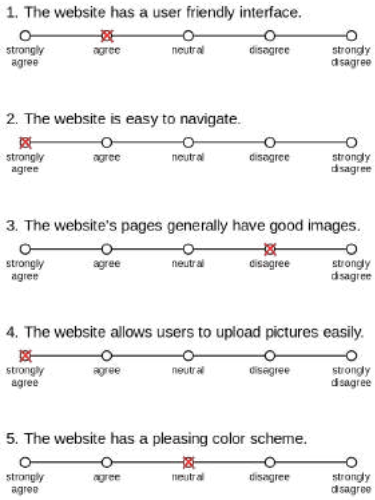
Questionnaire Design

- Open-Ended Questions – Free response text
  - Use opened ended when:
    - impossible to determine all possible answers in advance
    - the list of options would be unusably long
    - capturing numerical data (can always be grouped later)
    - qualitative aspects of user experience

Questionnaire Design

- Closed-Ended Questions – Predefined answers
  - Use when there are small number of possible answers
  - Using rating scales and ordinal data
  - Unipolar construct: 0 to extreme amount
    - Importance, usefulness…
    - Labels: “Not at all ...,” “Slightly ...,” “Moderately ...,” “Very ...,” and “Extremely ....” - shown to be semantically equidistant
  - Bipolar construct: Extreme negative to extreme positive
    - Ease of use (difficult to easy), visual appeal (unappealing to appealing)…
    - Labels: “Extremely ...,” “Moderately ...,” “Slightly ...,” “Neither ... nor ...,” “Slightly ...,” “Moderately ...,” and “Extremely ....”
  - Can have single choice, multiple choice, ranking and rating
  - All these choices have implications on how you analyse the data and what kind of statistics you can complete

Likert Scale



Questionnaire Design

- Measurement error: deviation of answers from true values on the measures
  - Can come from respondent
    - Lack of motivation
    - Lack of comprehension
    - Lies
  - Can come from instrument
    - Wording / design flaws
    - Technical / interaction flaws
  - No opportunities for clarification
  - Usually only deploy once
    - Can't revise halfway through, then consider all results when respondents have answered slightly different questions

Questionnaire Design – Bias

- Each question must be carefully checked for bias
  - A bias introduced in one question can even affect subsequent
- Look at 5 types

Questionnaire Design – Bias

- Satisficing
  - Surveys require focus, motivation, and/or cognitive load
    - Respondents don't put in the effort: fail to follow >=1 of Tourangeau's 4 cognitive steps
  - Weak: might pick answer that's OK but not optimal
    Strong: pick answer completely randomly
  - Avoid options such as "no opinion" or "n/a" if you want to force a choice
    - Can avoid by offering even number of possible responses on scale
    - But what if opinion is genuinely in the middle?

Questionnaire Design – Bias

- Satisficing (cont)
  - Avoid repetitive questions that use the same scale ('straight-lining')
  - Avoid overly long questionnaires
  - Communicate importance of survey to increase motivation
  - 'Trap' questions

Questionnaire Design – Bias

- Acquiescence Bias refers to respondent being more likely to agree in the presence of a binary design
  - Avoid yes/no questions, phrase questions neutrally
- Social Desirability Bias refers to responses given because the respondent thinks it will be looked upon favourably
  - On a scale of 1-5, how important is climate change research?
- Response Order Bias refers to respondents being more likely to choose responses at the beginning or end
  - Primacy/recency effects
  - Categorical answers should be presented in a random order
  - Rating scales (positive->negative, negative->positive) can be counterbalanced

Questionnaire Design – Bias

- Question Order Bias
  - Same as experimental design, order effects should be mitigated where order effects may occur
  - You might keep demographic questions in the same order at the beginning of a survey, but randomise the remaining questions
  - In cases where randomisation doesn't make sense consider organising by:
    - Starting with more general questions and finishing with specific questions
    - Starting with easy questions and finishing with difficult question − limit amount of dropout
    - Start with most important questions − limit impact of dropout

Questions To Avoid

- Broad Questions
  - Provide noisy data and confuse respondents
- Leading questions
  - Influence respondents and add noise/bias to data
- Double-barrelled questions
  - Conflate multiple constructs and make clear conclusions impossible
- Recall Questions
  - Self report from the past is inaccurate and noisy
- Prediction Questions
  - Self prediction is very susceptible to bias

Don’t reinvent the wheel…

- Some questionnaires exist that are widely tested, validated, and accepted as standards
  - SUS (System Usability Scale)
    - One of most commonly used. 10 questions (efficiency, satisfaction...), yielding single score
  - NASA TLX (Task Load Index)
  - UEQ (User Experience Questionnaire)
  - etc

Review and Testing

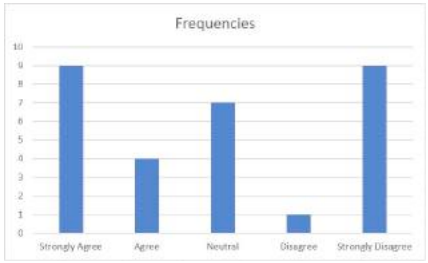
- Running a pilot to determine realistic completion times and check any bugs/ configuration issues/etc are fixed before public launch
  - Minor tweaks to survey configuration can mean all the data collected thus far is invalid or incomparable to data collected moving forward

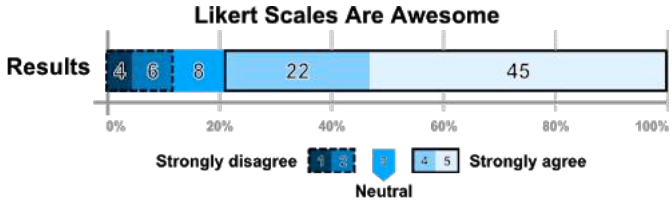
Analysis and Reporting

- Can learn certain findings e.g. user attitudes from interviews, but with surveys can get statistically reliable metrics
- Quantitative and qualitative analysis possible
  - More on qualitative analysis next week
  - Statistical analysis the following week

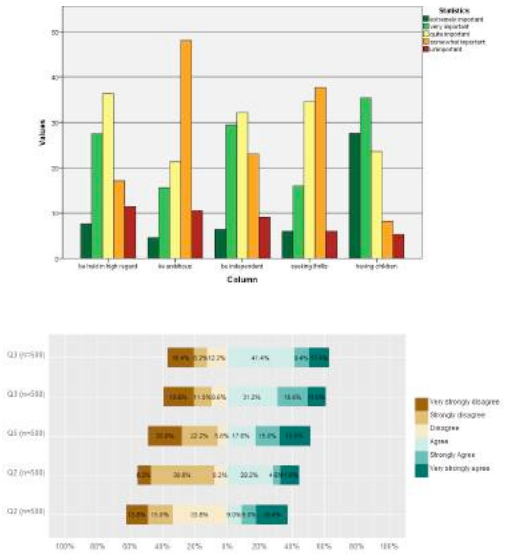
Likert Scale – Present Results

Various options





Likert Scale – Multiple Results



Experience Sampling

- ‘Ecological momentary assessment’
- Regularly fill out several brief questionnaires
  - Daily, or maybe several times a day
  - At specific times, and/or by responding to alerts
- Sampling regularly, don’t know participant circumstances. So limit burden
  - Closed-ended, fast, not dozens of questions
- Ask about current activities and feelings
  - NOT recall: reduce cognitive biases of memory-based self-report methods

For reference: https://ieeexplore.ieee.org/document/1203750

Intercept Survey

- Deploy survey while person using technology
  - e.g. Pop-up in an app while in use
  - Real-time data capture − minimise issue of imperfect recall
  - Can be triggered by particular behaviour of interest
  - Might be very annoying
  - Design the timing − maybe not while using feature, but some time soon after
    - Balance precision of recall / getting in users' way

Using Surveys with Other Methods

- Combine larger and smaller-scale
  - Use a survey
    - Captures high level information from a broad group of users
  - … along with a focus group, lab study, or interview study
    - Captures detailed information from a smaller group of users
  - Is data representative or just anecdotal? What are reasons for large-scale trends?
- Also keep record of all user interactions
  - Then can compare survey responses when user has done X vs when user has done Y − create user 'groups'
    - E.g. Impressions of product − compare users who skipped tutorial vs those who completed it

Focus Groups

- Involve bringing together a group of participants for a group discussion
  - Can be of various sizes. 3-6 people common in HCI. Sometimes 6-12, or more
  - Can involve structured activities or be run as a group interview
- Video/audio is recorded and analysed using qualitative methods
- Develop a protocol/script

Focus Groups

- Can be efficient way of getting many viewpoints
  - e.g. 4 hour-long groups of 6 people
- Participants can debate issues among themselves
- Maybe people don't want to raise issues in a group they would be more likely to tell you if they were interviewed on their own
- Beware someone monopolising conversation
- Split into separate groups − not everyone at once
  - More time for each person to contribute
  - 'Groupthink' and conformity can bias members. You can spot these potential patterns if separate groups give opposite consensuses

Focus Groups - Experience Prototypes

- Prototypes can range in fidelity, but give developers, designers, and potential end users hands on experience with a prototype
  - Focus on creating an experience, especially during the early stages of design when a fully functional prototype doesn't exist

Focus Groups: Keep/Lose/Change

- In groups, facilitating positive, negative, and creative feedback can be achieved using a "Keep/ Lose/Change" activity
- After experiencing a prototype, application, or demonstration, focus groups are asked:
  - What would you keep in the interface?
  - What would you lose or remove from the interface?
  - What would you keep but change?
- Often works best with large printouts of interface views that participants can mark up and annotate with post-its as a group

Ethnography, Interviews, and Qualitative Methods

Lecture 9-10

Interactive Systems

Ethnography

- Understanding and describing social and cultural scene from insider's perspective
- Roots in anthropology
  - Studies of non-Western cultures
  - Attempt to develop deep understandings of unfamiliar civilisations
  - Local people as pursue daily lives in own communities
- Fieldwork
  - Dispassionate observer insufficient - engage directly with people in everyday lives
  - Firsthand encounters to gain understanding
  - Deeply embedded perspective to get insights otherwise impossible

Ethnography

- The highly influential Chicago School innovated the use of ethnography as a method for understanding everyday life
  - E.g urban sociology
  - Local, maybe familiar settings
- Still based on immersion in context / community / culture
  - Understand how people go about everyday business
  - How organised
  - Standards and norms

Ethnographic perspectives

- Focus on predictable, daily patterns of human thought and behaviour
- Interpret observed behaviour in culturally relevant context
- Allow multiple interpretations of data/reality
- Open minded approach − allows exploration of rich sources of data not mapped out in research design
- Ethnographer is a human instrument
  - Senses, thoughts, feelings; very sensitive and perceptive data gathering tool
- Bias
  - All researchers have bias; make it explicit
  - e.g. choice of what to study is biased. Controlled, can focus and limit research effort; uncontrolled can undermine research quality

Fieldwork

- Being there, observe, ask insightful questions
- Interviewing: “ethnographer's most important data gathering technique”
  - Explain and put into context what see/experience
  - Study every word for subcultural connotations
- Document what you’ve seen and heard
- Notepads, audio, video, photo, survey
  - Analysis at various stages - field notes, reports
- Information gathered can be subjective and misleading
  - Cross-check, compare, triangulate before use as a basis of knowledge
- Classical ethnographies might spend 6 months to 2 years on fieldwork. Maybe 2 weeks every few months

Ethnography in HCI

- Combination of observation, interviews, participation
- Computer use as communication / collaboration
  - Use in existing groups (work, education), or purely virtual (forums, communities)
  - Norms and dynamics that might be important to study
- How systems are used
  - How design affects the way they’re used

Ethnography in HCI

- Just understanding
  - Groups, communication, new technologies…
- Different stages of design cycle
  - Early stages to gain deep understanding of system requirements
  - Later stages to gain deep understanding how a product is being used (maybe in a particular setting, or by a particular group), so can redesign to better support users
  - Study combination of range of technologies in a particular setting

Example − designing a new system for unfamiliar domain

- Need to understand system requirements
- Can be rooted in context of how target users work and interact
  - Organisational concerns
  - Work practices
  - People's values
  - Types of interactions between people
- Don't assume users are 'just like you'

Example − designing a new system for unfamiliar domain

- Could we use surveys / interviews to learn about requirements instead?
  - Maybe − that's certainly easier/cheaper!
  - If early stages and this is unfamiliar area, don't know what to ask
  - People's descriptions of what they do are often inaccurate
    - Poor at explaining
    - Misremember
    - They don't realise what they do
    - Bias (e.g. socially acceptable answers)

Example − designing a new system for unfamiliar domain

- Can learn more through conducting site visits
  - Potentially for days/weeks
  - Observe
  - Interview
  - 'Shadow' them
  - As start to understand how they work and what they need, can begin listing requirements & designing
    - Discuss with potential users − for approval or to correct misperceptions
    - Try with different users, possibly in different setting

Participant Observation for Design Inspiration

Ethnography in online communities

- Analysed collaborative play in World of Warcraft
  - Authors performed a lot of gameplay – active participants
  - Wanted to know what players were experiencing
  - Make recommendations to improve

For reference: https://dl.acm.org/citation.cfm?id=1180898

Observation Techniques

- Observation
  - Passive observation of everyday activities without active participation or intervention
  - Maybe not integrating into any community − just watching public spaces
- Participant Observation
  - Or ‘participatory observation’. Combines participation in the lives of those being studied with appropriate professional distance

Participant Observation

- Complete participant
  - Become part of community as much as possible
  - May take years
  - Risk losing ability to be detached − “going native”
  - Covert observation − don’t tell community you’re a researcher. Ethically challenging!
- Complete observer
  - Opposite end of scale. Don’t interact directly
- Can integrate quicker into 'own' culture − already an ‘insider’
  - But if too familiar, can take events for granted and leave important data unrecorded

Ethnography Challenges

- Requires a lot of skills
  - Skills in conversation
  - Data interpretation
  - What to pay attention to
  - Whom to talk to
  - Reconcile contradictory data
- Expensive
  - Often used in 3 contexts
    - Users not well understood
    - Tasks not well understood
    - Safety-critical systems

Interview Techniques

- Structured Interviews
  - Each participant answers same questions
  - Verbal approximation of questionnaire
  - Maybe appropriate when have explicit research goals
- Semi Structured Interviews
  - Each participant answers the same questions, but additional questions and follow-up questions can be added as needed
- Unstructured Interviews
  - Interview may have little or no set structure
  - Could be tool for early evaluation, where don't have firm idea of specific research questions

Designing an Interview

- Survey Questions
  - Designed to elicit a broad picture of the participant's experience
  - Good for building rapport and establishing scope
- Specific Questions
  - Designed to gather feedback on specific categories, attributes, and themes
- Open and Close Ended Questions
  - Balance of structured and unstructured responses

Designing an Interview

- Many issues familiar from survey lecture
- e.g. recall bias − if asking about past behaviour, do it soon after
- Close-ended questions
  - Do you have previous experience with VR?
- Open-ended questions
  - Tell me about your previous experience with VR
- Might be better to do as interview rather than survey
  - Probably longer open-ended answers
  - Can ask follow-up questions
  - But much more time consuming

Running an Interview

- Respect for the context the interviewee is coming from
- Respect for the interviewee
- Strategies
  - Be honest, be yourself
  - Focus on learning from participants
  - Be perceptive, know when to press and when to let go
  - Understand silence and use it
  - Being a good interviewer is a skill you can develop with experience

Key Actors

- In ethnographic setting, “some people are more articulate and culturally sensitive than others”
  - You will also find some users respond better to your ideas, provide more useful feedback, and act as “star users”
- Balance star users or key actors with the dataset
- Over-reliance can be dangerous
  - Cross-check with others to ensure they’re providing reliable information

Ethics Checklist

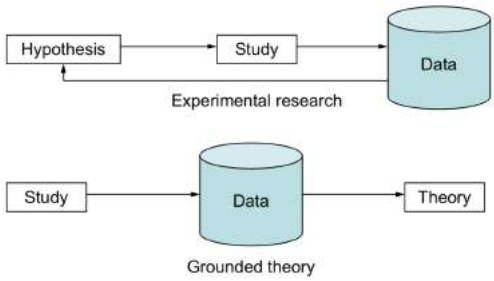
1. Participants were not exposed to any risks greater than those encountered in their normal working life.
2. The experimental materials were paper-based, or comprised software running on standard hardware
3. All participants explicitly stated that they agreed to take part, and that their data could be used in the project
4. No incentives were offered to the participants.
5. No information about the evaluation or materials was intentionally withheld from the participants.
6. No participant was under the age of 16.
7. No participant has an impairment that may limit their understanding or communication.
8. Neither I nor my supervisor is in a position of authority or influence over any of the participants.
9. All participants were informed that they could withdraw at any time. All participants have the right to withdraw at any time during the investigation.
10. All participants have been informed of my contact details.
11. The evaluation was discussed with all the participants at the end of the session, and all participants had the opportunity to ask questions.
12. All the data collected from the participants is stored in an anonymous form. All participant data (hard-copy and soft-copy) should be stored securely, and in anonymous form.

Analysing Qualitative Data

- Analysing data from e.g.
  - interviews
  - focus groups
  - open-ended questionnaire responses
- Transcribe any audio data
- Familiarise yourself with all data
- Deductive coding
  - A-priori codes search for; clear pre-existing questions
- Inductive coding
  - Find all the themes in the data

Analysing Qualitative Data

- Inductive approaches
  - Thematic analysis
  - Grounded theory
    - No preconceived theories; open mind
    - Theory eventually 'emerges' from the process



Analysing Qualitative Data

- Qualitative Coding – loosely separated 'stages'
  - Stage 1. Open coding
    - Identify distinct pieces of information; assign open code
    - In-vivo coding: use participants' own words to define codes
    - Size / scope of pieces determined by researcher's interpretive process
  - Stage 2. Axial coding
    - Organise open codes into set of concepts / categories
    - Think about relationships between concepts
    - Don't need to all be same level of specificity, or need even numbers of codes assigned
  - Stage 3. Selective coding
    - In grounded theory, combing concepts into main theory
    - Re-code original transcripts using new concept framework
  - Can verify with multiple coders at various stages

Example transcript

- From focus group on viewing media while travelling
  - participants numbered; interviewer is 0

0: OK, so first question. When you travel on an airplane, do you typically watch movies either on the provided screens on the seat back or on a personal device. And if so, which ones?
1: I watch them on the seat
2: Last time I had a flight that had one of those, which was before smartphones were really prolific, I just used that. I think now if I had like my tablet with me, I would prefer to just watch it on that to be honest.
0: You prefer the tablet
2: Purely because I get to choose what's on there.
0: ok
3: Same reason for me as well. I would normally choose my own personal device because you can view what you want.
0: So in the case of a tablet, would you be satisfied with that experience? With like the screen size, the level of privacy you get?
3: The screen size is good but there are issues with privacy in that sense everyone else can see what you're watching and also the other thing is that tablets can be a bit unwieldy on an airplane.
2: The logistics of it is a pain.
3: It's easier with a smartphone, based on all the same privacy issues.
0: ok
4: Yeah, I normally use a laptop. The screen is bigger but again, just issues with the placement of it.

Reporting Results From Qualitative Analysis

- If e.g. a thematic analysis uncovers 5 main themes
  - Five subsections, where explain each issue
  - "You can provide participant quotes" [p12]
    - Can relate to a user summary table, where 1 row per user and info provided on age, level of experience, job…
- Discuss overall findings
  - Put in context of related work – reinforce other findings, contradict, expand scope, consider different factors…?
  - Might lead to implications for future designs

Statistics for User Studies

Lecture 11-12

Interactive Systems

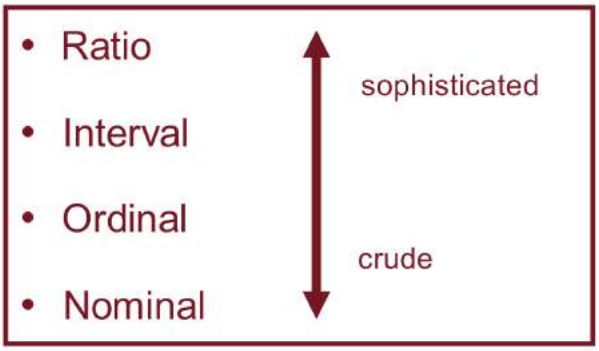
AE1-3 and 1-4

- AE1-4 this week in lab
  - Contact other team to participate
  - Someone else will contact you
  - Let me know if people don't respond

Analysing Data from User Studies

- Providing "descriptive statistics" is the bare minimum
  - Average, distribution, standard deviation
- Making claims, inferring causal relationships, in terms of a hypothesis test
  - Have you shown that your product is "better" than existing approaches?

Measurement Scales



As judged from the types of computations possible with each measurement

Measurement Scales

- Nominal / categorical
  - Labels or names
    - These could be numbers, but can't do computations with them, e.g. randomly generated IDs
- Ordinal
  - Can put the values in a ranking, but not equally spaced
  - e.g. ordered list of favourite films
  - Can do > or < comparisons, but not valid to calculate means

Measurement Scales

- Interval
  - Equal distances between adjacent values, but no absolute zero
  - Can compute mean
  - e.g. Celsius scale
    - Can take mean value of week's temperature, but strictly speaking 20°C is not "twice as hot" as 10°
  - Likert scale
    - Sometimes treated as Ordinal or Interval. Important to know which if you want to compute means. Treating as Interval OK if options are equally spaced and centred at neutral value

Measurement Scales

- Ratio
  - Do have absolute zero
  - Support many calculations
    - add, divide, mean, standard deviation…
  - e.g. time, distance, counts of events

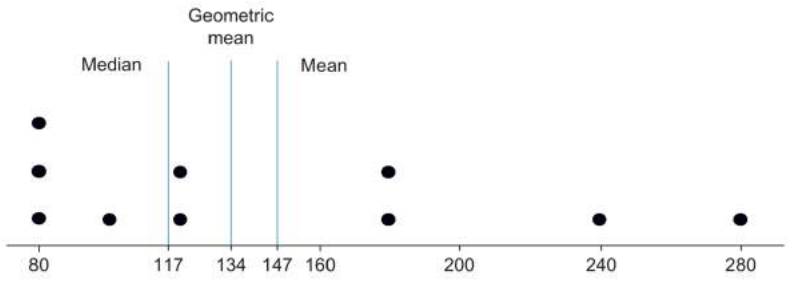
Evaluations and Measurement

- Before you can do anything, you need to plan well and measure the right dependent variable by collecting the right kind of data
- Types of Data
  - Think about data in terms of qualitative and quantitative
  - Think about quantitative data in a spectrum from continuous to discrete

Descriptive Statistics

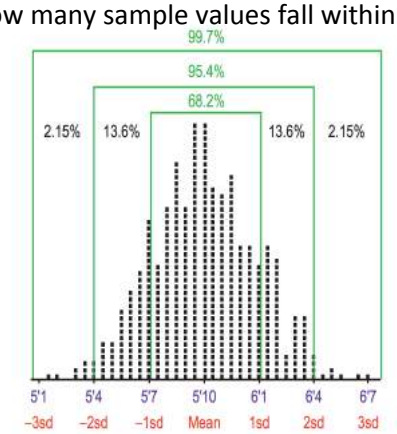
- Measures of Central Tendency: Mean, Median, and Mode
  - The mean is simple to calculate, but also provide little (or potentially misleading) information
    - Typically only useful if normally distributed data
  - The median may differ significantly from the mean, and this can insight into the "shape" of the data
- Standard Deviation describes the spread of the data
  - Estimate of average difference of values from the mean
- Plotting distributions tells you much more than simple values

Descriptive Statistics



Standard Deviation and Normal Distribution

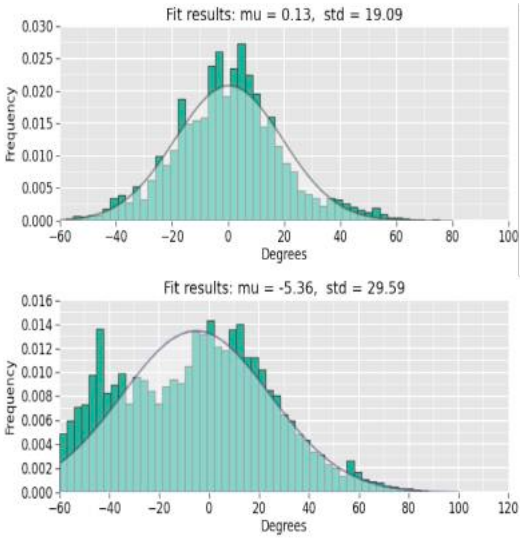
Empirical Rule: how many sample values fall within x Std Devs of mean



z-score: number of standard deviations from mean

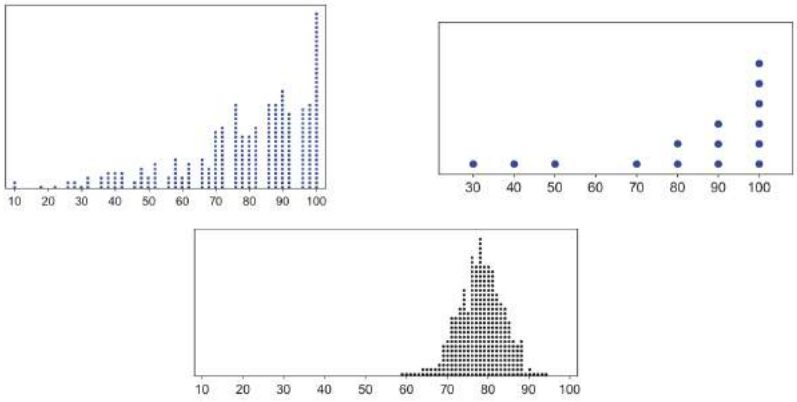
Standard Deviation for Evaluation Data

With human participants, the data is typically not normal distributed



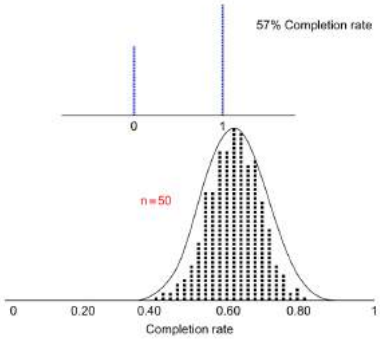
Central Limit Theorem

- As the sample size approaches infinity, the **distribution of sample means** will follow a normal distribution regardless of how parent population distributed



Central Limit Theorem

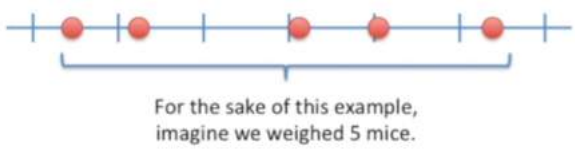
- As the sample size approaches infinity…
  - But often say > 30 (even smaller for interval data)
- Even applies to binary data:



Central Limit Theorem: implications

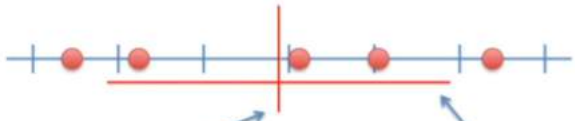
- Many statistical hypothesis tests (e.g. t-test) assume normal distribution of data
  - If data non-normally distributed (e.g. skewed), will these tests be invalid?
  - If sample size is large enough, CLT tells us that the distribution of sample means approximate a normal distribution
  - And so, we can use these hypothesis tests!

Standard Error



https://www.youtube.com/watch?v=A82brFpdr9g

Standard Error



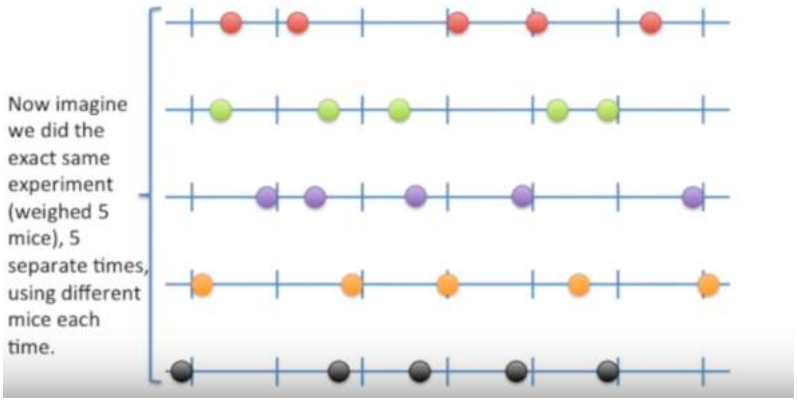
This is the average (or mean) of the values we measured.

This is the **standard deviation** on both sides of the mean.

It quantifies of how much the data are spread out.

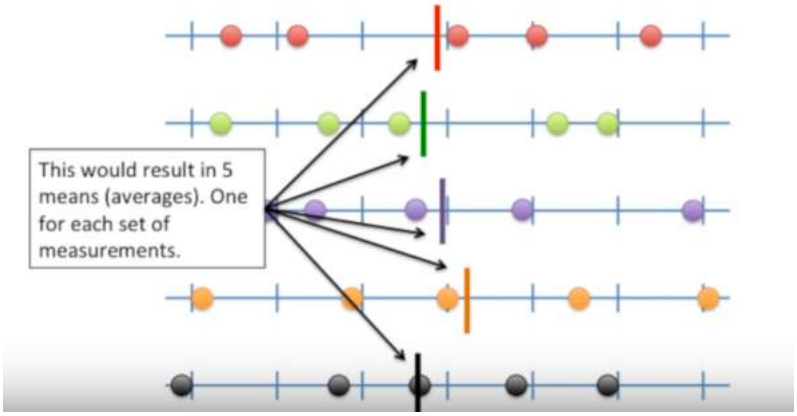
https://www.youtube.com/watch?v=A82brFpdr9g

Standard Error



https://www.youtube.com/watch?v=A82brFpdr9g

Standard Error



https://www.youtube.com/watch?v=A82brFpdr9g

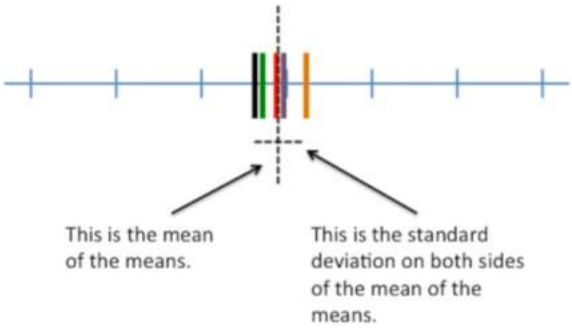
Standard Error

Here's what it would look like if we plotted all 5 means on the same number line.

https://www.youtube.com/watch?v=A82brFpdr9g

Standard Error

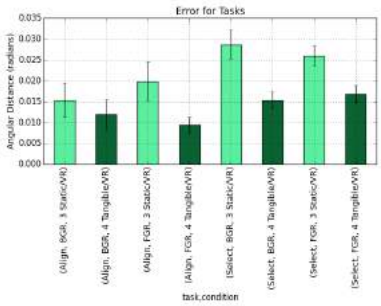
The standard deviation of the means is called The Standard Error.



https://www.youtube.com/watch?v=A82brFpdr9g

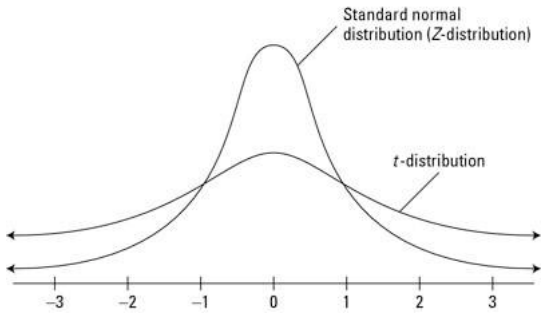
Standard Error

- But we don't want to run multiple experiments/take multiple samples
  - How we model error in means from samples given we do not know the mean of the population?
  - **Estimate** as: sample std dev / sqrt(sample size)



t-distribution

- Can't know from our experiments about distributions, means, st devs of *population*, only our sample
- Student's t-distribution, t-scores rather than z-scores

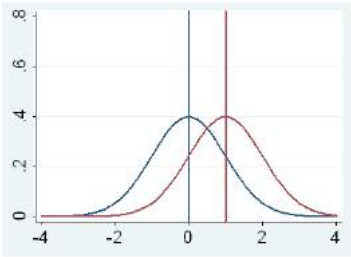


Hypothesis Testing

- Null Hypothesis
  - Rejecting the null Hypothesis
- Why is research done this way?
  - Very hard to prove something scientifically
  - Much easier to disprove
- Consider the following two statements:
  - Every software project has errors
  - Software projects never have errors

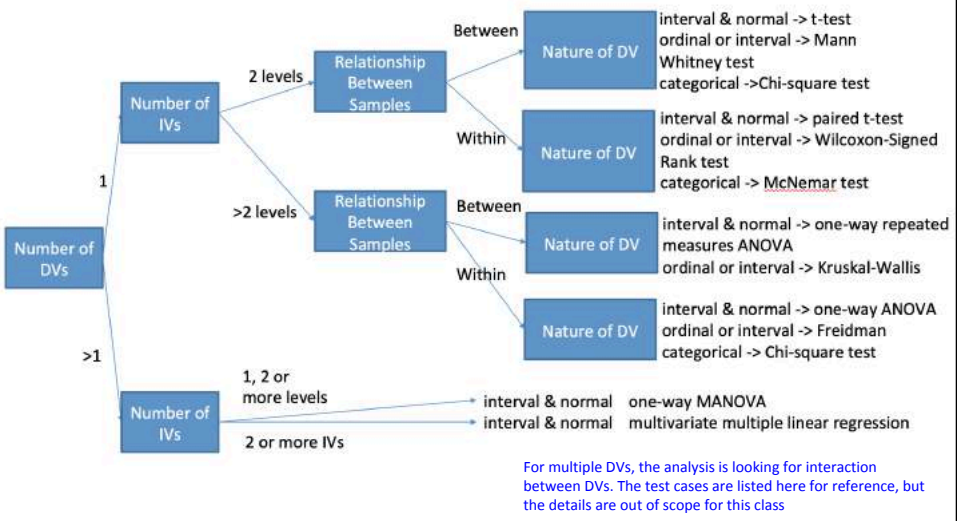
Hypothesis Testing

- Probably looking for sufficient evidence (instead of definitive proof)
- What are stats tests testing? **How likely is it that two samples come from the same distribution?**
- Also interested in:
  - How confident are we that they're different?
  - By how much are they different?



Hypothesis Testing

- Consider an example comparing a mouse to a trackpad
  - Null hypothesis: There is no difference between user performance in using these two input devices for an object selection task
    - If we reject the null hypothesis, we can analyse the data to present results arguing for where differences occur and what gains this may have for interaction



t-test and paired t-test

- The t-test was developed by chemist William Gosset ("Student") working at Guinness in 1908, quantitatively measuring the quality of beers
- Assumptions
  - Data follows a normal distribution
  - Data drawn from interval/ratio data
- Can be completed on dependent (within subjects) datasets with *paired t-test*, or independent datasets

Friedman and Wilcoxon Tests

- Friedman
  - Participants rate the quality of n different wines
    - Null Hypothesis: There is no difference between the wines
- Wilcoxon
  - Used for pairwise comparison, can provide results describing which wines are rated *significantly* better than others
  - Signed comparison: better or worse?
- Tests for a difference in related samples (within subjects)
- Used for ordinal data or interval data that is not normally distributed

Mann-Whitney and Kruskall-Wallis Tests

- Kruskall-Wallis is like the Friedman but for independent samples (between subjects)
- Mann-Whitney is like the Wilcoxon but for independent samples
- Consider our same scenario, participants rate n number of wines
  - Null hypothesis: Participants are unable to discern the difference between wines
  - Kruskall-Wallis test will tell us if there is variance across participants (for example by grouping participants by experience with wine tasting) and Mann-Whitney will provide pairwise comparisons to compare each group

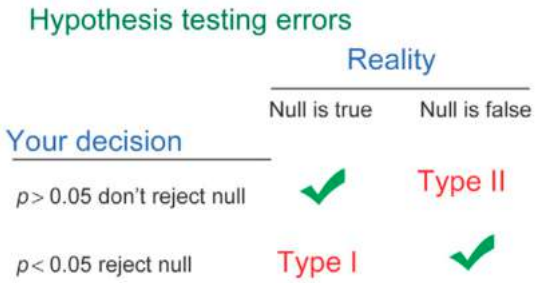
How to Present Statistical Results

- Each test produces a *p* value (the probability that the samples come from the same distribution) and a test statistic
  - Can choose a target for p; often say p<0.05 means statistically significant
  - Test statistics are interpreted differently for each test
- Most test would also be presented with an effect size
  - This value ranges from 0 to 1 and describes how "visible" the effect is

For reference: https://www.statisticsdonewrong.com/

Errors in Statistical Testing

- Type 1: False Positive
- Type 2: False Negative



Theories of Design
&
Modelling Interaction

Lecture 13-14

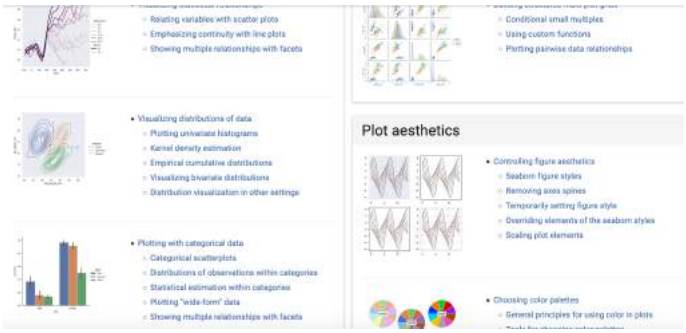
Interactive Systems

Labs and Exercises

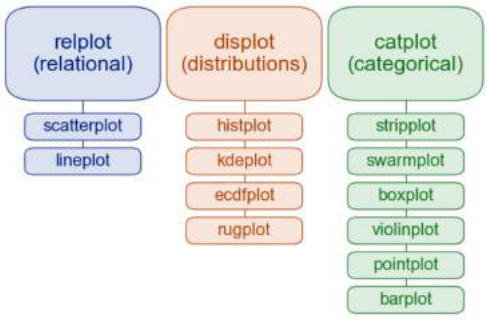
- Assessed Exercise 1 nearly done!
  - Marks within a couple of weeks
- No more team work!
  - Online labs still managed through Teams chat
- This week's lab − unassessed
  - VR acceptability dataset, analysis, stats, charts
- Assessed Exercise 2 released start week 9
  - Individual exercise
  - Typing dataset, analysis, stats, charts
  - Deadline 3rd December 18:30

Creating charts

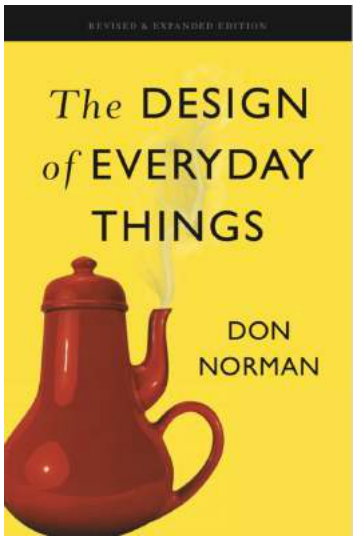
- Tutorial links on moodle for Matplotlib and Seaborn
- Plotly an option too – maybe more complicated
- Often good strategy to browse galleries



Seaborn



- If matplotlib “tries to make easy things easy and hard things possible”, seaborn tries to make a well-defined set of hard things easy too
  - Not always ‘well-defined’ for your needs


REVISED & EXPANDED EDITION
The DESIGN of EVERYDAY THINGS
DON NORMAN


Light switches should be as follows:


No dogs
except guide dogs
PULL
PULL

Affordance

- Possible interactions between people and environment
- *The relationship* between physical object and person
  - Not a *property* of an object
  - Objects convey important info about how people could interact with them
  - Presence of affordance jointly determined by object's properties and person with capabilities that determine how it could be used
    - A chair affords sitting
    - A chair affords lifting to some people

Anti-Affordance

- Prevention of interaction
- To be effective, affordances and anti-affordances must be discoverable
  - If it can't be perceived, need to signal its presence with a *signifier*

Signifier

- Communicate behaviour
- Image, text, sound…
  - Make an affordance apparent
- Deliberate
  - E.g. Labels
- Emergent
  - Paths worn onto ground
  - Queues of people

Shoogle: Physical Affordances in a Digital Interaction

Keys in a pocket. The user carries the phone in a pocket while walking. Motion from the gait of the user is sensed by the accelerometers. As messages arrive, objects begin jangling around in the user's pocket, in a manner similar to loose change or keys.

Knowing What To Do

- How can you make unfamiliar situations feel familiar?
- Knowledge in the world
  - Perceived affordances, controls & their actions
- Knowledge in our heads
  - Experience, conceptual models, constraints

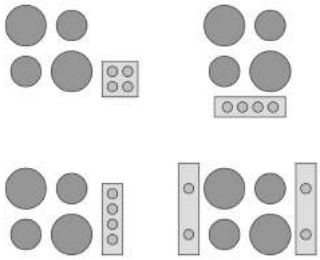
Constraints

- Physical Constraints – rely on properties of the physical world
  - e.g. can only insert the correct way; USB-A, bank cards
- Cultural Constraints – rely on socially learned behaviours
  - Moodle relies on roles that make sense to us because we know how a course is run

Constraints

- Semantic Constraints – rely on intrinsic meaning
  - When added all items to buy, look for control for checkout screen
- Logical Constraints – rely on trial and reasoning
  - e.g. An online form won't submit. Even if it doesn't highlight required fields, we can scan through and see if we left one empty − that one's probably the problem

Logical constraint problem?



Constraints

- Semantic Constraints – rely on intrinsic meaning
  - When added all items to buy, look for control for checkout screen
- Logical Constraints – rely on trial and reasoning
  - e.g. An online form won't submit. Even if it doesn't highlight required fields, we can scan through and see if we left one empty − that one's probably the problem
- Imposing these constraints prevents error and guide users towards correct, desired, or useful behaviour

Guiding Interaction with Constraints

- Forcing Functions – Preventing action until certain requirements are met
  - Interlocks – Requiring actions to occur in sequence
    - Web app that doesn't offer you functionality until you're logged in
  - Lock Ins – Keeps an action active, preventing action from stopping
    - Gmail checks if an attachment is attached before sending email
  - Lock Outs – Prevents an action from occurring (typically in safety context)
    - Operators of x-ray machine cannot enter dangerous values

Where should forcing functions be?

Cash machine software varies significantly. In some cases, the card must be inserted in the machine for the entirety of the interaction. In some cases, the card is only inserted for the beginning of interaction.

Cash machines also vary as to when your PIN must be entered. Some machines require this before any options can be selected, some require this after an option requiring security is selected.

Where **should** forcing functions be implemented?

A: At the beginning of all interactions
B: Distributed throughout interactions at the point where security is needed
C: Some other approach?

Forcing Functions and Usability

- Balance error prevention with frustration

Convention

- Design consistency is virtuous
  - lessons learned from one system transfer to others
- If can't put knowledge in the design, put it into a cultural constraint
- Standardisation
  - Maybe a last resort; when no other solution seems possible, at least do everything the same way

Convention vs Progress

- People don't like change
  - New learning is required
  - Which is 'better' design is irrelevant - the change is upsetting
  - Better to be consistent if new way is only slightly better than old?
  - If change to a new system, everyone has to change - mixed systems confusing
- Standards simplify life, but can hinder future development

State Machines Can Be Used to Model Interaction

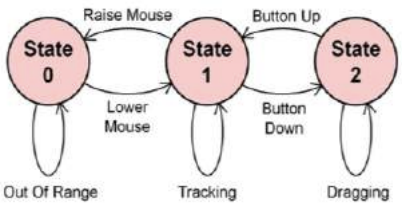


FIGURE 7.9

Buxton's three-state model for graphical input.

(Adapted from Buxton, 1990)

Fitts' Law

- Model to predict the speed of people's movements
  - One of few 'Laws' in HCI
  - Very widely used, proven to hold on many forms of interaction devices
- Paul Fitts, 1954
  - Psychologist, work predates HCI

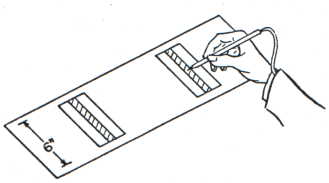
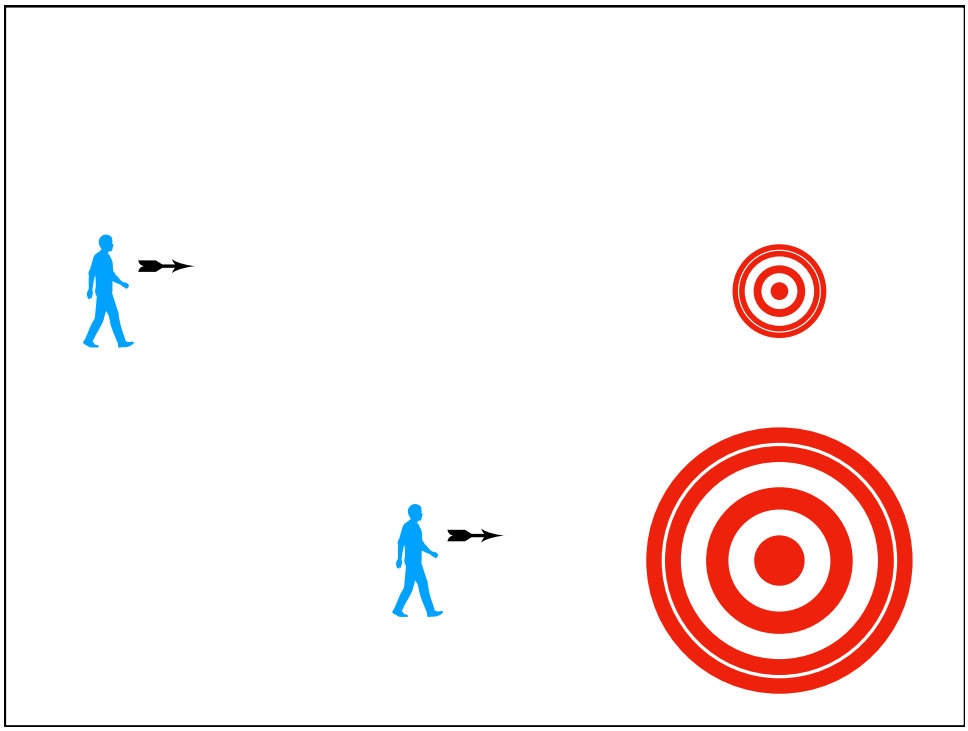


FIG. 1. Reciprocal tapping apparatus. The task was to hit the center plate in each group alternately without touching either side (error) plate.

Fitts' Law in HCI

- Most commonly used form adapted for HCI by Scott MacKenzie
- Ease to acquire a target a function of size of and distance to target

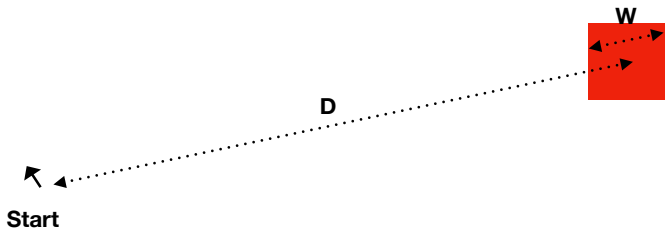


Fitts' Law Equation

- $MT = a + b * log_2(\frac{D}{W} + 1)$
- MT - time to complete a movement
- D - distance from start point to target
- W - width of target (how accurate you need to be on arrival)
- a & b - constants determined by cognition, hand-eye coordination - often different for different device type
  - Out of our control in terms of designing on-screen positioning within interfaces

Fitts' Law Equation

- $MT = a + b * log_2(\frac{D}{W} + 1)$

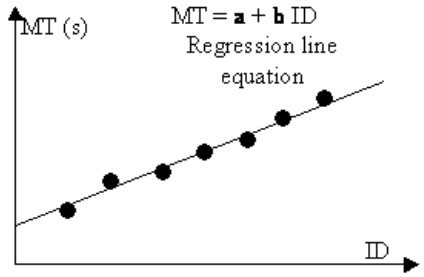


- W is measured along the axis of motion
- Pragmatically, often measured using the minimum of width, height

Fitts' Law Equation

- $MT = a + b * log_2(\frac{D}{W} + 1)$
  - $MT = a + b * ID$
- $ID = log_2(\frac{D}{W} + 1)$ ID = index of task difficulty (bits)
- $IP = \frac{ID}{MT}$ IP = index of performance, or throughput (bits per sec)

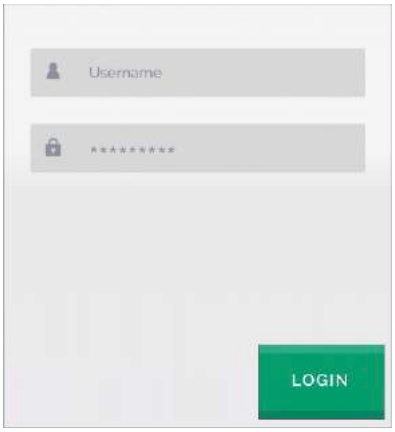
Linear regression model

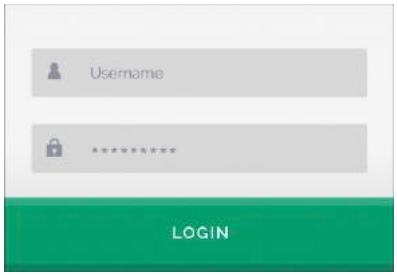


Purposes

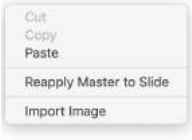
- Many hundreds of studies have confirmed Fitts' Law holds with different devices, input methods (e.g. mouse vs trackpad vs touchscreen)
- Fitts' Law can be used in
  - Predicting movement time (if a and b are known)
  - Comparing two devices (by comparing their IP values)
  - Guiding design choices

Design Implications





Design Implications



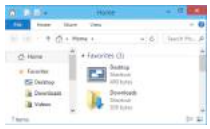
Easiest places to reach

- The easiest place: where we are right now!
  - Right-click menus: pop-up in place
- Screen edges - can't overshoot, so don't have to be accurate
  - Effectively a target of infinite width in a pointer-based interface
  - Corners especially good

Content bound to edges

- MacOS menus always bound to top

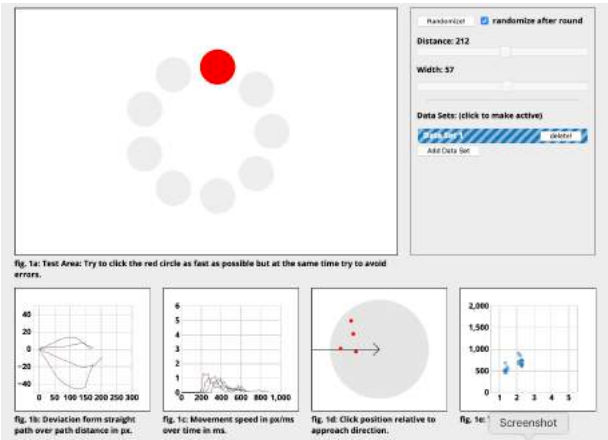
Content bound to corners



- PC - close window with ‘X’ in corner - easier to reach
  - But is that the desired behaviour?

Interactive Demo

- Try it out: http://simonwallner.at/ext/fitts/



Large-Scale Trials

Lecture 15-16

Interactive Systems

A/B Testing

- Many names, same idea
  - A/B testing, web experiments, control/treatment, randomised experimental design, controlled experiment, split testing...
- Randomly split traffic among different app versions
  - A/Control: usually current live version
  - B/Treatment: new idea
- Collect metrics and analyse

A/B Testing

- Previous lectures have talked through design principles
  - But still hard to accurately predict if successful
  - Often small features will have a surprisingly big effect
- *However compelling the message, however great the copy, however strong the sales argument… the way a page is designed will have a dramatic impact on conversion rates, for better or for worse.*
  - http://www.alistapart.com/articles/designcancripple/

Some design elements that can make a significant difference to page performance

- The position and prominence of the main heading
- The number of columns used on the page
- The number of visual elements competing for attention
- The amount of 'white space' on a page, giving the content space to 'breathe'
- The age, sex and appearance of someone in a photo
- The position and colour of the primary call to action
- Position on the page of testimonials, if used
- Whether linked elements are shown as text or as images

Amazon Shopping Cart Recommendations

- Add an item to your shopping cart at a website
  - Most sites show the cart
- At Amazon, Greg Linden had the idea of showing recommendations based on cart items
- Evaluation
  - Pro: cross-sell more items (increase average basket size)
  - Con: distract people from checking out (reduce conversion)
- HiPPO (Highest Paid Person's Opinion) was: stop the project
- Simple experiment was run, wildly successful, and the rest is history

For reference: R. Kohavi et al. Practical Guide to Controlled Experiments on Web. Proc. ACM KDD 2007

A/B at Microsoft

- "*We have an unprecedented opportunity to run A/B tests with online users and innovate more quickly based on actual user response. Microsoft needs to shift the culture from planning the exact features to planning a set of possible features, and letting customers guide us.*"
  - Ray Ozzie, Chief Software Architect at Microsoft

MSN Real Estate

- "Find a house" widget variations
- Overall Evaluation Criterion(OEC): Revenue to Microsoft generated every time a user clicks search/find button

A

B

• Does A or B win?

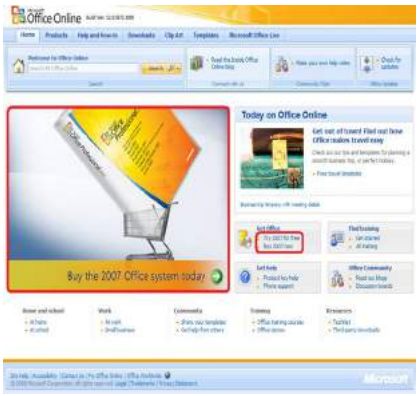
MSN Real Estate

- A was 8.5% better
- Since this is the #1 monetisation, it effectively raised revenues significantly
- Actual experiment had 6 variants.
  If you're going to experiment, try more variants, especially if they're easy to implement

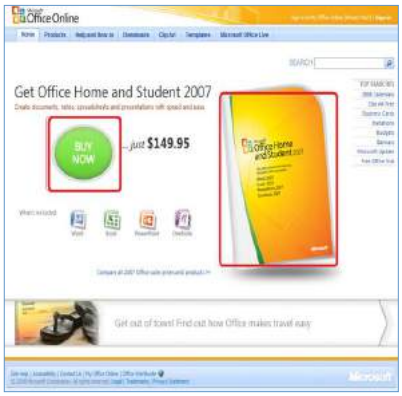
Office Online

OEC: Clicks on revenue generating links (red below)

A



B



• Does A or B win?

Office Online

- B was 64% worse
- Doesn't tell us why one is more successful
  - Interviews, qualitative methods
- But main point: it is hard to assess the value of ideas
  - Get the data by experimenting because data beats intuition

Ramp-Up

- To detect an effect, you need to expose a certain number of users to the treatment (based on power calculations)
- Fastest way to achieve that exposure is to run equal-probability variants
  - e.g., 50/50% for A/B
- But don't start an experiment at 50/50% from the beginning: that's too much risk
- Ramp-up over a short period
  - Start an experiment at 0.1%
  - Do some simple analyses to make sure no egregious problems can be detected
  - Ramp-up to a larger percentage, and repeat until 50%

Advantages of A/B testing

- It tests for *causal* relationships, not just correlations
- It reduces the effect of external factors
  - e.g. history/seasonality impact both A and B in the same way
- Overcome poor intuition, especially with novel ideas
  - All too often: the less data, the stronger the opinions
  - So get the data through experimentation

Problems with A/B testing

- Organisation has to agree on OEC (Overall Evaluation Criterion)
  - This is hard, but it provides a clear direction and project alignment
- Quantitative metrics may not explain *why* a treatment is better or worse
  - And therefore help designers to solve problems or know where to go in the next design iteration

Problems with A/B testing

- Primacy effect
  - Changing app or site may degrade the user experience (temporarily) even if the new design is better
- Multiple experiments
  - Statistical variance increases, making it harder to get statistically significant results
- Consistency and contamination
  - Assignment to A or B usually cookie-based, but people may use multiple machines or erase cookies
- Be careful to do proper randomisation!

Large-Scale Mobile HCI Studies

- Mobile HCI studies take many forms:
  - e.g. text entry, gestures, AR, usage studies, privacy…
- How do you perform studies?
- Quantitative analysis
  - e.g. time taken to complete task / error rates…
- Qualitative analysis
  - e.g. interviews, ethnography, opinions of experience…

Into the wild

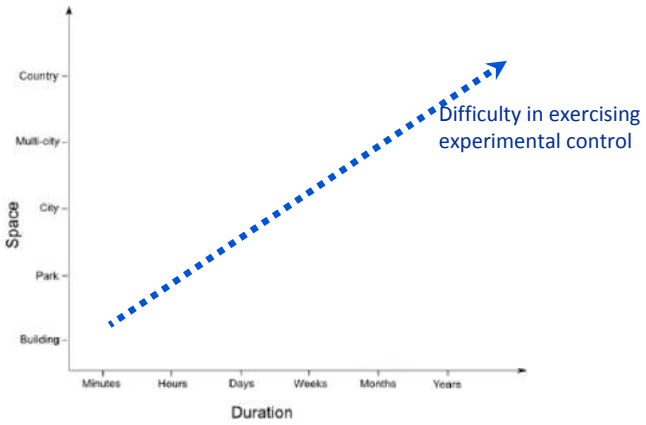
- Early / 'traditional' experiments all done in the lab
- Easy to observe, control, eliminate confounding variables
- But maybe unrepresentative of technology's eventual intended context of use
- More recent studies often performed in more realistic settings: 'in the wild'

Into the wild

- What forms of evaluation are used in the wild?
- Even if outside a lab, often still direct observation
  - Videos, interviews
- Often still using evaluator-supplied hardware



Challenges of Space & Duration of Trials



Research via app stores

- Put software you want to study on app stores
- Participants using their own devices
  - Already experts with hardware; no training
  - No extra device to carry; already have with them always
  - No fixed end date
- Potentially very large numbers of users, from all over the world
  - Chosen to use the app; more representative?

Research via app stores

- Not all good news?
  - Don't meet users
  - Can't directly observe users
  - Can we still gather qualitative data?
  - Internal vs external validity
  - Additional ethical challenges?

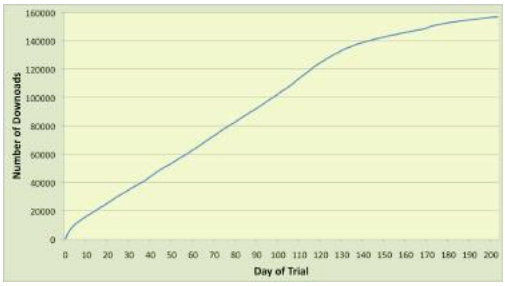
Example - Hungry Yoshi

Game using Wi-Fi access points as game resource

For reference: https://doi.org/10.1007/978-3-642-12654-3_13

Example - Hungry Yoshi

- Investigating use of app stores in running mobile HCI trials
- > 300,000 downloads; new users quite steadily acquired
- Only worked for around 2 years, as iOS updates blocked functionality



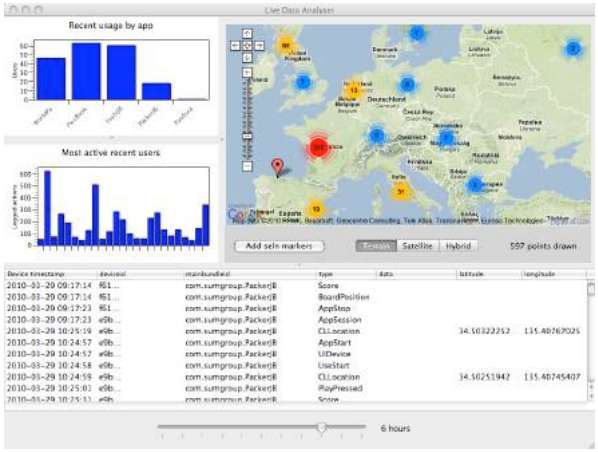
Example - Hungry Yoshi

- Global user base
- Only have locations from those users who agree to supply it

Data logging

- How can we tell how people are using our software around the world?
- Need to record ("log") info on use while apps are running
- Hungry Yoshi used logging framework with phone and server parts
- Capture user interactions (button taps, screen changes), device/sensor data (accelerometer, Wi-Fi connections)
- Write to locally-stored files on device
- Opportunistic uploading over Internet

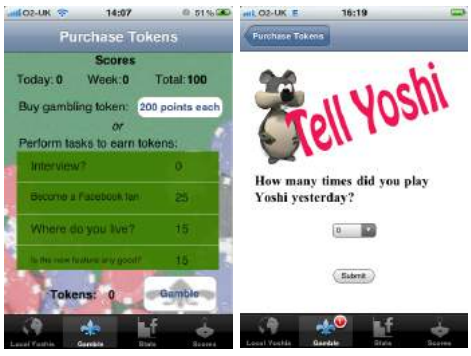
Data visualisation



- SGVis - live analysis of usage

Qualitative evaluation

- Would it be possible at a distance in this type of trial?
- Questionnaires
  - Answers selected from lists or type into textbox
  - Other types of 'task', e.g. become a Facebook friend
  - 19% of players responded
- Server-side, so instant updates
- Paid telephone interviews



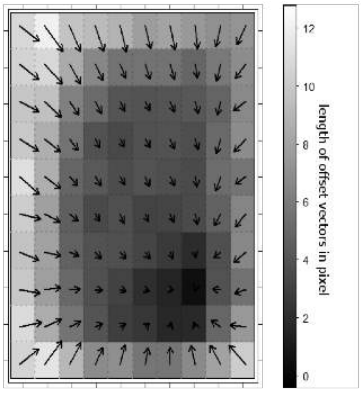
Example: Hit It!

- Android game: touch objects on the screen
- 91,731 installations
- 120,626,225 touch events
- Looked at error rates for different sizes and locations

For reference: https://dl.acm.org/doi/10.1145/2037373.2037395

Example: Hit It!

- Found that touch positions are skewed
- Could create function that shifts touch input to compensate
- Updated game to use compensation
  - Error reduced by 7.8%



Large Scale Trials: Difficulties

- Verification of user info
  - Are people telling the truth or just ticking random buttons?
    - About age, gender, opinions...?
- Trial software installed on large variety of devices
  - Especially on Android
  - Differences in OS, CPU, screen sizes…
  - All could be confounding variables

Large Scale Trials: Difficulties

- Collecting qualitative data is difficult
  - Very short questionnaire answers
  - Phone / Skype calls
    - Most users probably don't speak your language
    - Difficult to arrange - time zones
- Mass of quantitative data; harder to get qualitative
  - Could see ***what*** was happening, but not ***why***

Hybrid Methodology

- Hybrid approach: combining ‘mass participation’ and local deployments
- Concurrent large scale and small scale trials
- App released to general public and local users recruited via poster adverts
- Some aspects of trial best suited to each group
- More solid ethical practice

Example: Predictor

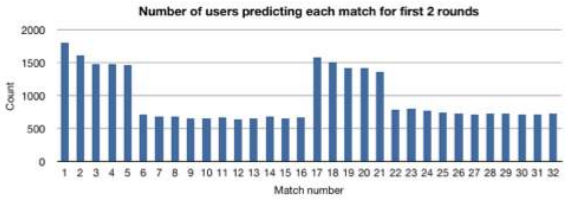
- World Cup Predictor app
- Released 1 week before football World Cup
  - 11 locally-recruited users
  - 10,806 through app store

For reference: https://dl.acm.org/doi/10.1145/2207676.2208588

Hybrid Trials: Benefits

- 1. Use the small to explain the large
- 2. Use the large to verify the small
- 3. More solid ethical practice

Use the Small to Explain the Large



Use the Large to Verify the Small

- If a system is trialled among small group of locally-recruited participants...
  - Do results generalise to population at large?
  - 'Outlier' users
    - Do we have user(s) showing unusual behaviour?
    - Could skew the results of the study

Use the Large to Verify the Small

- Experimenter effects
  - Subtle conscious or unconscious cues an experimenter gives participants
  - Could affect users' performance in the trial
  - Less likely in large-scale trial?
    - User interaction with evaluators generally far lower

Use the Large to Verify the Small

- Looking at one feature of app: head-to-head challenge
- Local users:
  - Head-to-head uptake: **73%**
  - Average number of H2Hs by those using: **5.2**
- Global users:
  - Head-to-head uptake: **0.8%**
  - Average number of H2Hs by those using: **1.5**
- **Local trial alone would have led to misleading results**

Ethical Challenges

- Capturing a lot of information on people
- Never meet participants
- Numerous ethical issues made more challenging by this methodology
  - Informed consent
  - No opportunity to debrief participants
    - Can't tell the last time a user will launch the app

Terms & Conditions page

- Page often shown on first launch
- Provides information on the experiment
  - Who we are
  - What we're studying
  - What we're logging and why
- Often have to be explicitly agreed to before can use app
- Opt-out mechanisms
- Multiple languages

Hybrid Approach: Levels of Engagement

- All participants agree to same T&Cs
- But difference in confidence with which researchers have gained informed consent
- High percentages not reading T&Cs
- Ease of deception
- Inability to debrief

Hybrid Approach: Levels of Engagement

- Framework of levels of engagement
  - Local users: studied in detail
  - Remote users: looked at aggregate data
  - Types of questions asked
    - Ability to converse sensitively at a distance
- Compromise: getting useful info, but not exploiting users as resource just because they tick T&C box

Ethical Challenges

- Informed consent
  - Do people know what we're doing?
    - That the app is part of University research?
    - The purpose of the experiment?
    - What information is being recorded?
    - What we will do with this info?
    - How to opt out?

Terms & Conditions page

- State the purpose of the study - URL to project site
- All logging explained and must be explicitly agreed to before app usage
- Store / transmit data securely
- Email address opt out at any time on request - have all collected data destroyed
- Multiple languages

Do users read the T&Cs?

- Hungry Yoshi study: "Did you read the T&Cs and realise you were taking part in a research trial?"
  - In-app questionnaire:
    - 1,226 responses. Yes: 20% No: 80%
  - Telephone interviews:
    - 11 interviewees. Yes: 0 No:11
- Opening the full T&C document:
  - 75,818 agreed to T&Cs. 2% opened doc
  - Of 2%, nobody spent >60 sec reading the 842 words

Researching Ethics

- Interpreting existing guidelines to cover large scale mobile HCI
- Framework categorising trials based on participant 'risk'
- Advice for how to run each type of trial in ethically sound manner
- Experiments on new ethical procedures

Interpreting Existing Guidelines

- Human trials in Psychology: BPS & APA
  - Autonomy, Dignity, Self-Determination
  - Concern for Others' Welfare
  - Social Responsibility
  - Scientific Value, Integrity, Competence
- Internet-Mediated Research

General Guidelines

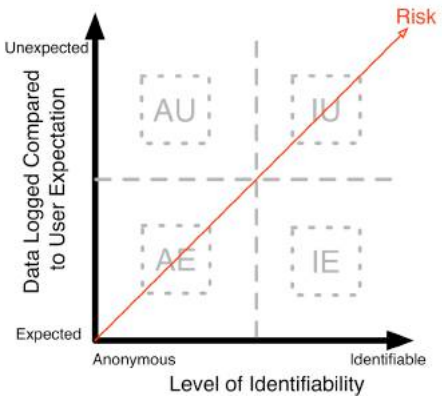
- Restrict age of users where stores allow
  - Graphics, icon sets, descriptive language
- Terms & Conditions in store description & in-app
- Historic log data not on externally-visible server
- Privacy-preserving data publishing techniques

Categorised Ethical Framework

- So many different forms of research - hard to make ‘one size fits all’ set of guidelines
- Identified 2 main dimensions of participant ‘risk’
  - Anonymous vs identifiable
  - User expectation of app’s data access
- Categorise existing trials on this framework

For reference: https://dl.acm.org/doi/10.1145/2470654.2466245

Categorised Ethical Framework



4 Quadrants

AE: Anonymous, Expected

AU: Unexpected, anonymous

IE: Identifiable, expected

IU: Identifiable, unexpected

Anonymous, Expected

- e.g: Aggregate download/usage stats
- e.g: Logging data that is integral to app usage, but cannot be used to identify user
- Generally low risk
- **Advice**
  - General guidelines sufficient
    - Terms & Conditions pages to explain research, etc

Anonymous, Unexpected

- e.g. a game looking at ‘unnecessary’ data: how many contacts you have, contents of media library...
- **Advice**
  - Pop-ups to gain explicit consent for each new data type captured
    - Mobile OSs now incorporate this

Identifiable, Expected

- e.g. location sharing apps, social media apps
- **Advice**
  - Provide functionality to browse data and delete specific parts
  - Effectively allowing ‘opt out’ at any time

Identifiable, Unexpected

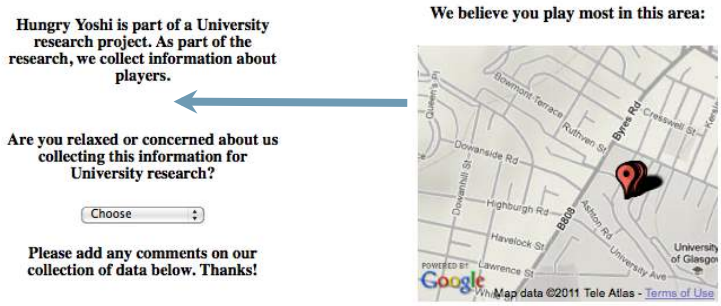
- e.g. a game looking at 'unnecessary' data that could identify a user: location
- Highest risk
- **Advice**
  - Actively interrupt users to show them examples of recorded data

Identifiable, Unexpected: Interruption

- T&C read rates suggest we need to find a better alternative to T&Cs
- Alternative idea, based on interruption
  - Visual representation of log data
  - Delayed presentation of information
  - Personalised with user's own data

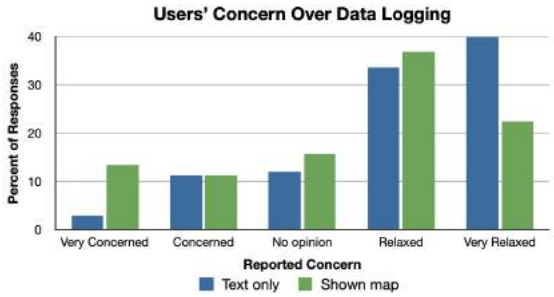
Interruption User Study

- 1007 users; between-groups design
  - Hash function on the device's unique ID to randomly assign to a condition



For reference: https://dl.acm.org/doi/10.1145/2639189.2639239

Concern Following Interruption



- Mann-Whitney U test: significant difference between conditions ($p<0.01$)

Further Results

- More concerned users stopped using the app around twice as quickly
- Difference of showing the map more pronounced for non-English speakers
- Also looked at age, gender: no significant differences

Comments received

- "*I have no problem! if this help to gather info I would participate in it*"
- "*I hope that your project goes well*".

Comments received

- "*I don't like that people knows that much about me*"
- "*U should not hav the right to do this!!!it's invading in peoples privacy :(now can u pls stop doing this!!*"

Discussion

- Seems that we should look beyond current common practice of T&Cs
- Majority of users seem relaxed
- A small number of concerned users, who we should be going further to support
  - Personalised visual representations of data
    - More users reported concern
    - Stopped using the application sooner

Discussion

- Can be extended to many forms of data
- Collect data only locally on device for a short period at start
- Interrupt user with visual depiction of her own data
- If she agrees to participation, upload all collected so far and keep logging
- If she disagrees, destroy collected data without it ever leaving the device
- Should be more engagement of users generally
  - Ethics as active area of research
  - Not just box to tick

Information Visualisation

Lecture 17-18

Interactive Systems

What is InfoVis?

- Visualise
  - *"to form a mental image or vision of..."*
    - Not just immediate perception, but fitting what's seen and interacted with into a mental model... and so updating that model
- Information Visualisation
  - *"The use of computer-supported, interactive, visual representations of data to amplify cognition"* [1]

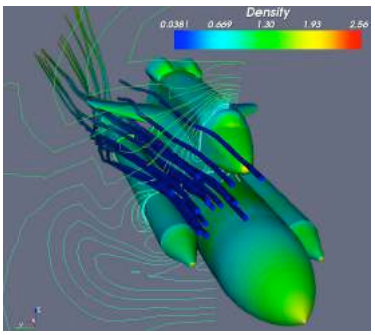
[1] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Readings in Information Visualization: Using Vision To Think. Morgan Kaufmann, San Francisco, California, 1999

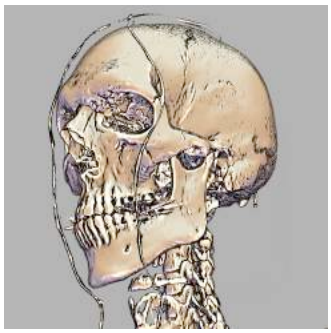
What is InfoVis?

- Forming mental model to gain insight
- 'Offloading' cognition
  - Reduce load on working memory
  - Using recognition instead of recall
  - e.g. think of doing long multiplication in your head
    - Much easier if you can write notes: 'workings'

What isn't Infovis?

- Scientific visualisation and cartography
  - Usually physical data about objects & spaces
  - Based on inherent or 'natural' dimensions





InfoVis is about visualising abstract data

X	Y	CUSTOMER_ID	PRODUCTTYPE	CURRENCY_IS	CUSTOMER_SI	YIELD	DAYS_TO_MAT	AMOUNT_CHF	FULLNAME	TRADE_DATE	INDEX
7.68573749	-8.551567	21	276	AUD	AAA	7.224	1002	250	AUD Eurobonds	1 Feb 1996	T00001
7.51999521	-8.6964572	21	276	AUD	AAA	7.266	1044	227	AUD Eurobonds	1 Feb 1996	T00002
7.49346291	-8.7835013	21	276	AUD	AAA	7.266	1044	227	AUD Eurobonds	1 Feb 1996	T00003
7.33578662	-9.168513	21	276	AUD	AAA	7.428	1068	227	AUD Eurobonds	1 Feb 1996	T00004
7.92332236	-4.4869782	53	156	CAD	AAA	6.169	981	176	CAD Eurobonds	1 Feb 1996	T00005
1.20632675	3.87790565	21	2	CHF	AAA	3.554	1841	100	CHF Domestic	1 Feb 1996	T00012
-0.3839602	3.44191014	55	2	CHF	AAA	3.89	2045	1300	CHF Domestic	1 Feb 1996	T00013
-1.5232496	5.42951462	96	2	CHF	BBB	3.796	2208	500	CHF Domestic	1 Feb 1996	T00014
-8.5371151	11.091578	10	2	CHF	AAA	3.956	2471	10000	CHF Domestic	1 Feb 1996	T00015
-4.21017	4.47227871	361	2	CHF	CCC	4.579	2581	100	CHF Domestic	1 Feb 1996	T00016
-4.2154588	4.69189146	361	2	CHF	CCC	4.204	2609	50	CHF Domestic	1 Feb 1996	T00017
...	...	...	...	...	...	...	...	...	...	...	...

- How best to present a data set?
  - Type of data?
    - Column types - Numerical? Ordinal? Dates?
  - Who's analysing it?
  - What are they looking for?
- Correlations, clusters, outliers...

InfoVis vs IR

- Information Retrieval: Absolutes
  - Maximum, average, exact query match
  - Formalised, suited for a command language
- InfoVis: Relatives
  - Overview, trends, patterns
  - Distributions and outliers, 'sense'
  - Difficult to formalise
  - Suited for interaction, browsing and exploration
    - ... built up over time via successive interactions...

Examples of InfoVis

- 'InfoVis' as subject been going about 30 years
- Earlier examples – e.g. London underground
- Harry Beck, 1933
- 'Circuit board' design
- Abstraction
  - Aid clarity
- Still used today

Key Principles

- Abstraction
  - replace many objects by one representative object
- Start with overview → Support zooming & filtering → Only then show details on demand
- Direct Manipulation
  - objects output on screen take input too
- Dynamic Queries
  - e.g. move a slider up and down, linked graph changes too
- Immediate Feedback
  - GUI interaction triggers response straight away

Key Principles

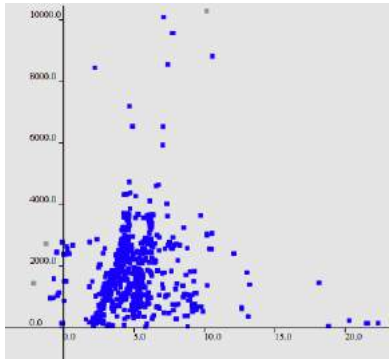
- Linking + Brushing
  - views linked so that selections match in all
- Focus+Context
  - show key objects in detail... but in the setting of the wider data set
- Animate transitions and change of focus
  - don't jump so harshly that context is lost
- Output is input
  - anything one can use to show data can be used to select data too
- Colour with care
  - be aware of colour blindness, (non)linear perception, visual overload

Representation

- Data encoded by:
  - Location
  - Size
  - Colour
  - Shape
  - Texture

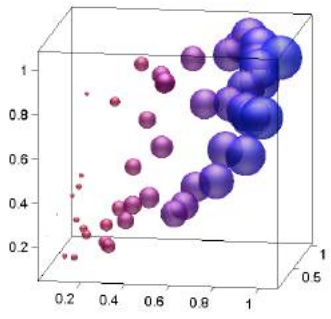
Encoding by Location

- Spatial location on display conveys value
- eg X-Y plots
  - scatterplot
- Can only encode 2 or 3 variables this way



Encoding by Size

- Size of points represent value
- Can run out of room very quickly
  - Occlusion: big points hide smaller ones
- What about negative values?



Encoding by Colour

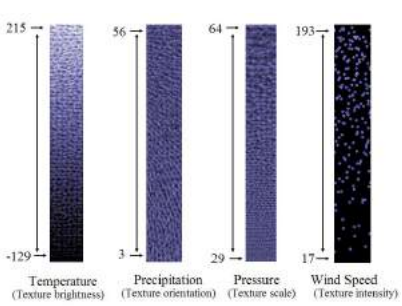
- Colour Scales: many to choose from
- Careful: RGB is a non-linear colour system
  - For example, (100,100,100) is not twice as bright as (50,50,50), and (100,100,0) + (0,100,100) ≠ (100,200,100)
  - Stick to a small simple palette: use highlight colours cautiously
  - Use a perceptually linear colour scale
- Minimum size at which visible

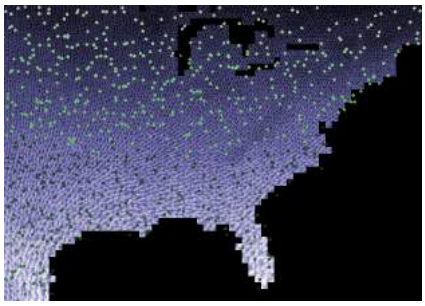
Encoding by Colour

- Perceptually linear colour scales
  - Arrays of RGB values scaled to better fit with average human perception
  - Colour at index 100 generally perceived as twice as bright as colour at index 50
  - More limited range: may have to avoid the many dark values at low indices

Encoding by Texture

- Colour scales are hard to read, but we can easily tell the difference between textures. eg. Tweed & Silk
- The finer the texture the closer we have to be to the graph to understand it.



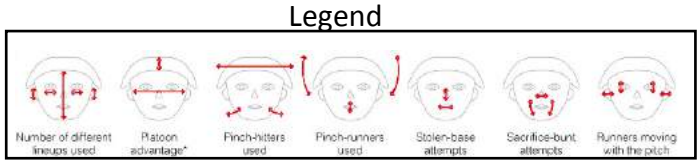


Ying Tang; Huamin Qu; Yingcai Wu; Hong Zhou; , "Natural Textures for Weather Data Visualization," *Information Visualization, 2006. IV 2006*

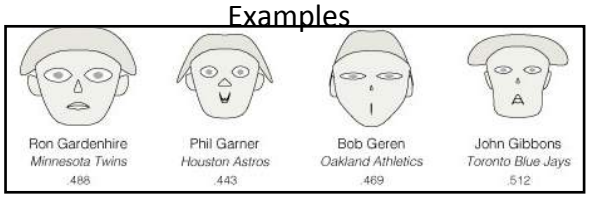
Encoding by Shape

- 'Glyphs' are used to visually represent multiple dimensions of data by combining them into a single pictorial representation.
- Most famous are Chernoff Faces
- Usually need the legend to understand them

Legend



Examples



Steve C. Wang; , "Visualizing Managerial Tendencies" *AAAS, 2008, New Techniques in the Evaluation and Prediction of Baseball Performance*

Ranking Visual Attributes

- Position
- Length
- Angle, Slope
- Size
- Colour

Increased accuracy for quantitative data

- Cleveland and McGill, 1984

Design guideline:
Map more important data attributes to more accurate visual attributes

Focus & Context

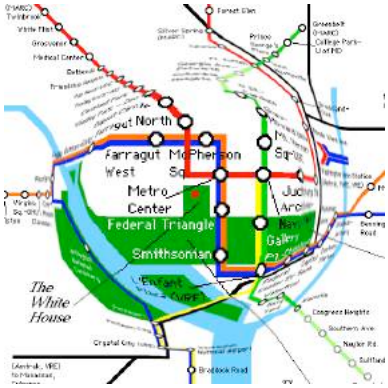
- Show detail as well as 'big picture'
- Maximise usage of available screen real estate
- Overview & detail
  - Area of detail and (usually smaller in screen size) overview covering larger area of data
- Distortions
  - e.g. blurring, fisheye

Focus & Context

- Overview & detail
  - Separate views
    - Can often interact with both

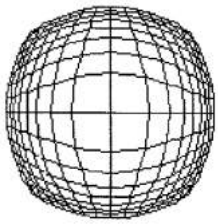
Focus & Context

- Distortions, e.g. ‘fisheye’
- Single view of the data
- Focus in high-detail, surroundings much less
- Normal vision involves perspective; things further away gradually get smaller
  - Fisheyes exaggerate the same effect
  - but still use smoothly increasing distortion



Focus & Context

- Distortions, eg 'fisheye' distortion
- An example metric:
- DOI (b|a) = API(b) - D(a,b)
  - DOI (b|a) : degree of interest in point b, given current focus a
  - API(b) : a priori importance of point b
  - D(a,b) : distance from a to b
- 'Information suppression' function
- General idea applicable in 1D, 2D...



Shneiderman's Taxonomy

- 7 types of data
  - 1D
  - Temporal
  - 2D
  - 3D
  - Trees / hierarchies
  - Graphs
  - Multidimensional (4+D)

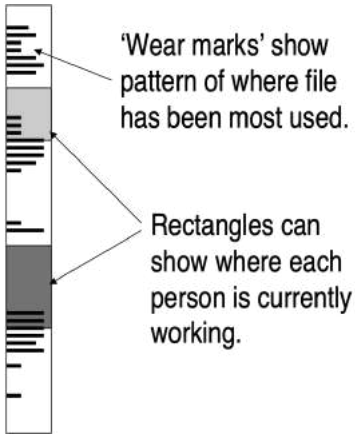
Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings of the Symposium on Visual Languages* (1996)

1D Data

- e.g. Single column of numbers or text
- List with scroll bar
  - People often only look at top of the list
  - can't easily see or move further down
  - But how to explore a long list/column?
- InfoVis techniques
  - Distortions

1D Data

- Edit Wear and Read Wear (Hill et al., 1992)
- Showing accumulated history of use
- Worn by use, like well-thumbed book pages, paths in grass & old stone steps
- Can extend well-known representations
  - e.g. fit into simple scrollbar
- Or fit into newer designs



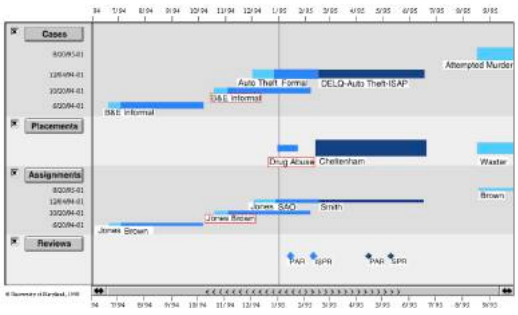
1D Data

- Experiment applying fisheye effect to 1D data
- Fisheye list faster to use than traditional for drag&drop tasks
- No difference for selection tasks
- Error rates same
- Users preferred fisheye

C. Wimmer, M. Tomitsch, T. Grechenig: "Evaluating the Usability of 1D Fisheye Lists" IHCI, 2009

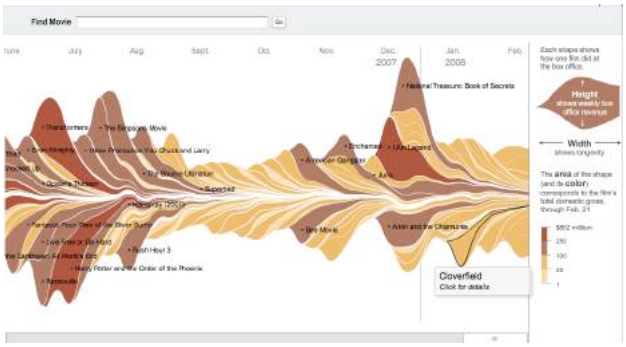
Temporal Data

- Very common: records, logs, databases
- Shneiderman distinguishes this from 1D
- Can ‘stack’ dimensions, sharing time axis
- Can use 1D techniques
  - e.g. fisheyes and distortions



Temporal Data

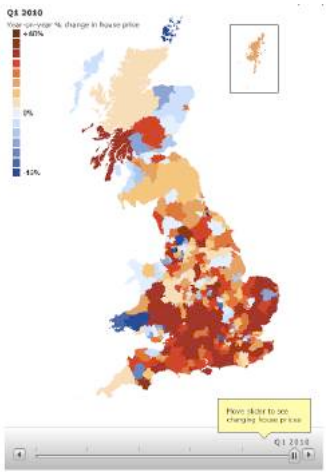
- The Ebb and Flow of Movies: 1986 — 2008
  http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html



- Interactive - mouse-over, scrollbar
  - but: overviews?
  - allow comparisons over time/patterns?

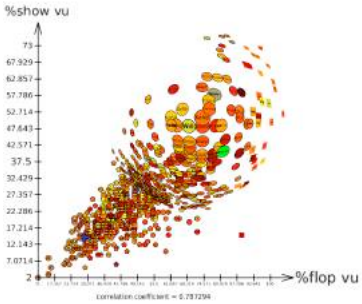
Temporal Data

- Timeline slider
  - 'playback'
  - Or can query specific times
- See pattern across country
- But hard to compare temporally distant data
- Again, no overview over time
  - Pattern detection more difficult
  - See everything - external cognition
  - Memory - internal cognition



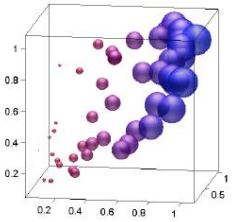
2D Data

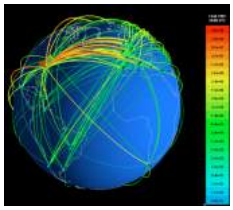
- Scatterplots - plot x vs y
- Maps: geographical data
- Fisheye, focus & context techniques again applicable



3D Data

- Appeal to the '3D is natural' idea
- Often think of the world as a 3D shape
  - But don't treat it that way
  - e.g. how wide is a city? How high?
  - e.g. can only see surfaces of most objects
- Often invites occlusion problems
- Nearby objects block distant ones
  - e.g. can only see half of a sphere
- 2D vs 3D: Design for what's easiest to use, not what looks cool

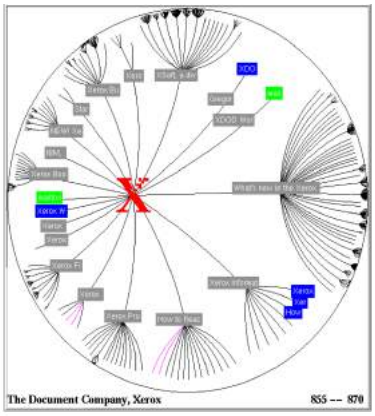




Global Network Traffic (AT&T)

Hierarchical data

- Trees are difficult to handle
- Basic problem is fan-out to many objects: can't show all the tree in detail all the time
- Hard to show many objects and lots of structure at the same time
- Can't avoid having to move around and explore
- Focus+context / distortion
  - e.g. hyperbolic tree
- Debate over glitz v. utility
- Experiments and design continue



Hierarchical data

- Tree maps: convert tree to rectangles
  - Area proportional to e.g. node size
  - Split space horizontally and vertically in turn

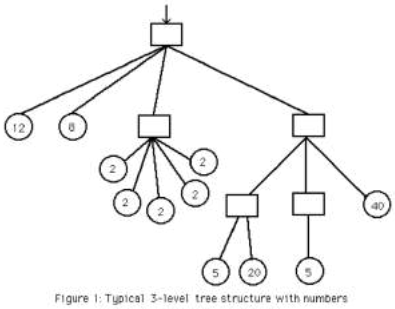


Figure 1: Typical 3-level tree structure with numbers indicating size of each leaf node

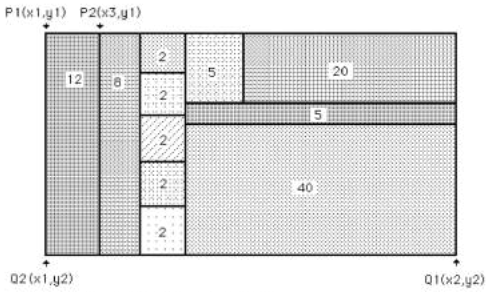
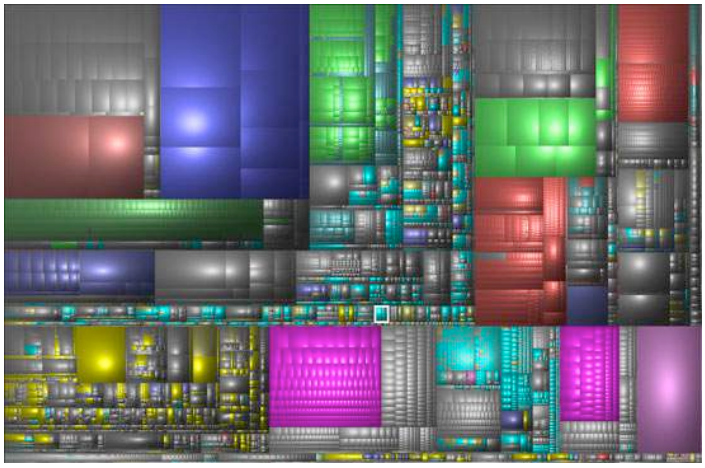


Figure 2: Tree-map of Figure 1

Hierarchical data



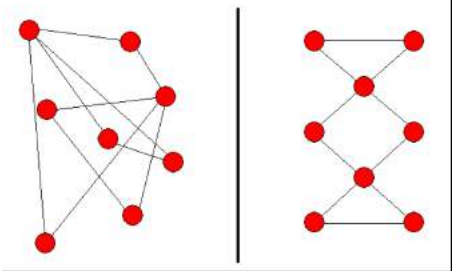
WinDirStat / Disk Inventory X

Hierarchical data

https://finviz.com/map.ashx

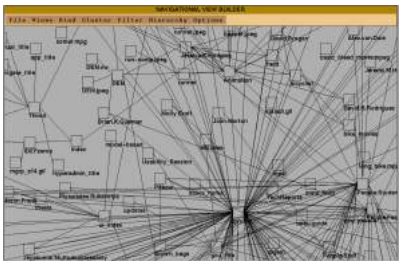
Graph Data

- Nodes and edges
- Aesthetics. 'Appealing' layout
  - Subjective?
  - Generally accepted desirable properties
    - Minimise edge crossings
    - Uniform edge lengths
    - Evenly space nodes
    - Symmetry
- 2 drawings of the same graph:



Graph Data

- Even more difficult to handle than trees
  - Links can go anywhere: may be no regular order/structure
- Optimisation algorithm
  - e.g. find positions that minimise edge length & crossings
  - Closely related to algorithms for multidimensional data (later)
- Sometimes it's good to just give up
  - reduce to simpler type e.g. convert to a tree: choose a root, lift it up, cut off or hide excess links

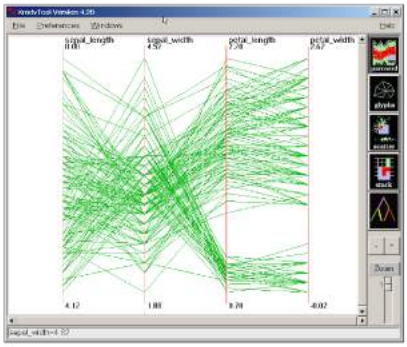


Multidimensional data

- Cleveland and McGill: humans best equipped to make judgements when data encoded by position
- But how to use position when dimensionality >3?
- Strategies for visualising multidimensional data
  - Non-orthogonal display of dimensions
    - e.g. Parallel Coordinates
  - Numerous Paired Combinations
    - e.g. Scatterplot Matrix
  - Dimensional reduction to create single scatterplot
    - e.g. Force-Directed Placement

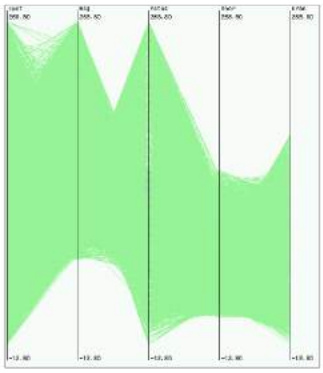
Parallel Coordinates

- Non-orthogonal display of dimensions
- Inselberg, 1985
- Each object a single polygonal line
  - Intersects each 'axis' at appropriate value
- See patterns, clusters...
  - 'Iris' data set: 150 objects, 3 natural clusters
- Good for correlations, if adjacent
  - Might need to re-arrange dimensions



Parallel Coordinates

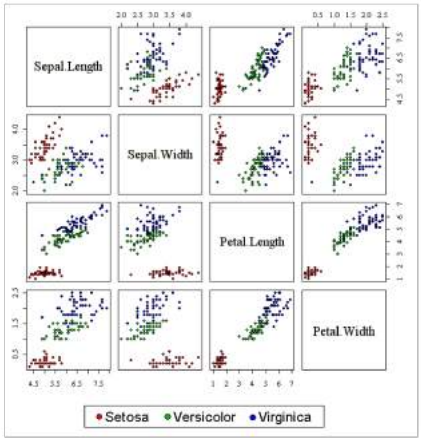
- Hard to follow a single object's line left to right
- Problem gets worse with bigger data sets
- Interactive controls can help
  - Mouse-over to highlight a single line



Parallel Coordinates

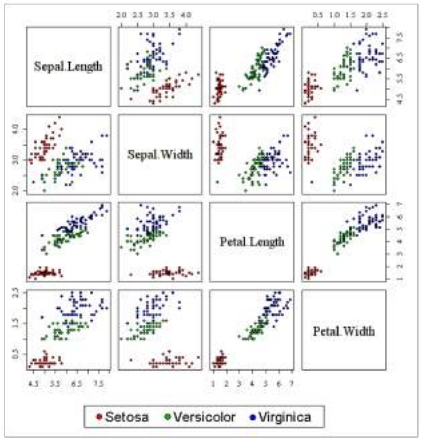
Scatterplot Matrix

- x-y scatterplots of every pair of dimensions
- Good for seeing correlations in pairs of dimensions
  - ...however they're positioned, unlike Parallel Coords



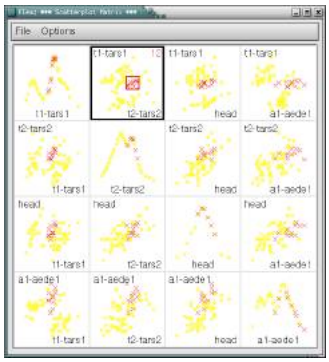
Scatterplot Matrix

- Duplication in grid: can just show 'triangle' either side of diagonal
- But still, screen space requirement rises quadratically with dimensionality
  - $\frac{1}{2}(d^2 - d)$ plots
  - (d = dimensionality)
- No overview of all the data



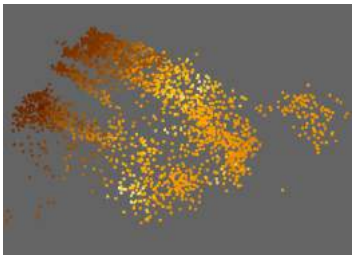
Scatterplot Matrix

- Again, interaction can make more powerful
- Brushing and Linking
  - Linking together multiple views, so that 'brushing' a selection in one view colours matching objects in other views



Single Plot Visualisation

- Create single scatterplot showing overall structure of data
- Compare objects: rows of the spreadsheet
  - Treat inter-object similarity as high-dimensional distance
  - Find a low-dimensional layout that retains as much of the relative distances between objects as possible
    - Similar objects close together in the layout, and dissimilar objs far apart
  - General approach often called dimensional reduction or multidimensional scaling (MDS)
    - Matrix methods (eg PCA), spring models...
- 'Reduce dimensionality'
  - to 2D / 3D layout
    - Seen already, 2D often preferable



Force-based models

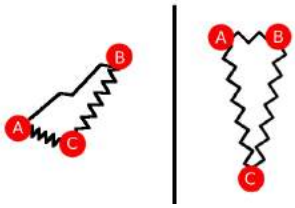
	X	Y	CUSTOMER_ID	PRODUCTTYPE_ID	CURRENCY_ISO	CUSTOMER_SEGMENT	YIELD	DAYS_TO_MATURITY	AMOUNT_CHF(000)	FULLNAME
A	7.68573749	-8.551567	21	276	AUD	AAA	7.224	1002	250	AUD Eurobonds Secondary
B	7.51999521	-8.6964572	21	276	AUD	AAA	7.266	1044	227	AUD Eurobonds Secondary
C	-4.2154588	4.69189146	361	2	CHF	CCC	4.204	2609	50	CHF Domestic Bond Secondary

- 'Spring model' to position objects
- Consider a spring between each pair of objects
  - Ideal relaxed length of spring proportional to difference between objects
  - i.e. from spreadsheet: objs A&B quite similar; C more different
    - AB is a short spring; AC and BC longer springs

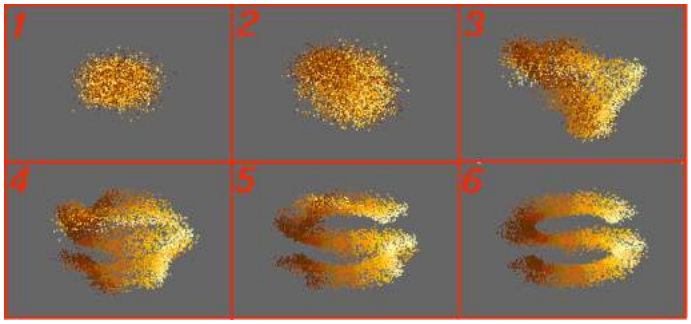
Force-based models

	X	Y	CUSTOMER_ID	PRODUCTTYPE_ID	CURRENCY_ISO	CUSTOMER_SEGMENT	YIELD	DAYS_TO_MATURITY	AMOUNT_CHF(000)	FULLNAME
A	7.68573749	-8.551567	21	276	AUD	AAA	7.224	1002	250	AUD Eurobonds Secondary
B	7.51999521	-8.6964572	21	276	AUD	AAA	7.266	1044	227	AUD Eurobonds Secondary
C	-4.2154588	4.69189146	361	2	CHF	CCC	4.204	2609	50	CHF Domestic Bond Secondary

- AB is a short spring; AC and BC longer springs
- Start from random positions (left image)
  - Some springs too stretched, others too squashed
- The springs then iteratively push and pul objects until the layout reaches equilibrium (right)



Spring Model Layout Process



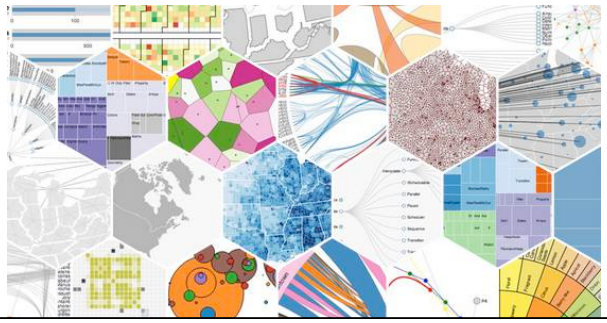
Snapshots from different stages of a spring model layout. Test set of "S"-shaped data set to help show what's happening. From random initial positions, structure slowly appears as similar (but far) objects pull closer to each other, and dissimilar (but close) ones push apart, in a simulated system of springs, forces and movement.

Spring Models: Strengths & Weaknesses

- Scatterplot layout positions show global relationships
  - Neighbours on layout are usually high-D neighbours
- All dimensions are combined in 2D layout
  - Not for exploring individual dimensions, unlike other techniques
- Can be very slow - often $O(N^3)$ overall time complexity
  - May be unable to lay out large/complex data sets, e.g. many millions

Examples

- Many examples in D3
  - JavaScript library for visualisation on the Web
    - HTML, SVG, DOM manipulation
  - https://bl.ocks.org/
  - Observable – Jupyter-style notebooks



Review and Exam Preparation

Lecture 19-20

Interactive Systems

IS Exam

- Wed 8th of December at 09:00 GMT. Online
- 90 minutes (**no longer double time allowed**)
- Changing format this year: two parts
  - Part 1: Multiple choice questions. **40 marks**
    - 20 questions, 4 options for each, only 1 correct
    - Negatively marked: Correct: **2** Incorrect: **-⅔** Blank: **0**
  - Part 2: 'Short essay'-style questions. **20 marks**
    - Each question will give guidance word count for answers
- NO programming, NO calculating stats

Week 2: General HCI and Visual Usability

- Assigned Reading: Visual Usability sections on Consistency, Hierarchy, Layout
- Iterative Cycle of Human-Centred Design
- HCI history
- Consistency helps interfaces feel familiar
  - Internal and External Consistency
- Visual Hierarchy is used to communicate the relative importance of elements
  - Characteristics of hierarchy
- Layout provides structure
  - Size, position, white space, scale, alignment
- Colour guides user gaze and attention
  - Convey hierarchy and relationships, colour models and types of contrast

Week 3: Human Perception and Capabilities

- Assigned Reading: MacKenzie book Chapter 2
- Scale of human actions
- Closed loop control
- Human sensing capabilities
- Human output capabilities
- Perception and cognition
- Human performance

Week 4: Lab Evaluation and Quantitative Methods

- Assigned Reading: MacKenzie book Chapter 5
- "Science is method. Everything else is commentary."
- Ethical Procedures
- Experimental Design
  - IV, DV, Control, Random, Confounding
  - Experimental protocol
- Running Evaluations
- Participants
- Order Effects and Counterbalancing

Week 5: Surveys, Focus Groups, and Qualitative Methods

- Assigned Reading: Ways of Knowing: Survey Research
- Survey Research
- Populations and Sampling
- Questionnaire Design and Bias
- Focus Groups and Experience Prototypes

Week 6: Ethnography, Interviews, and Qualitative Methods

- Research Methods in Human-Computer Interaction Ch 9
- Fieldwork
  - Participant Observation
    - Best practice
- Interview
  - Types of Interviews
  - Types of Questions
- Ethics
- Qualitative Analysis
  - Open, Axial, and Selective Coding

Week 7: Analysis Techniques and Statistics

- Assigned Reading: Fundamental Statistical Concepts
- Types of data
- Describing data
  - Overviews
  - Normal Distribution
- Central Limit Theorem
- Standard Error
- Hypothesis Testing
- Statistical Testing – common tests (from flowchart)
  - What are they, which/when to use; not how to calculate
- Statistical Errors

Week 8: Theories of HCI and Models of Interaction

- Assigned Reading: The Design of Everyday Things Ch 1 & 4
- Affordance
  - vs signifiers
- Constraints
  - Physical, Cultural, Semantic, Logical
- Forcing Functions
- Model of Graphical Input
- State Machine Models
- Fitts' Law

Week 9: Large Scale Studies

- No assigned reading
- A/B testing
- Scaling up Mobile HCI trials
  - 'Hybrid' trial methods
- Ethical concerns
  - Categorised framework
  - Mitigating methods

Week 10: Information Visualisation

- Assigned Reading: Spence book Chapter 3
- Definitions & Principles
- Data Representation
- Shneiderman's 7 types of data
- InfoVis for multidimensional data

2018 IS Exam

- Remember, exam changing this year. This had three 20-mark questions. You'll only get one 20-mark question in this style.(Plus 40 marks from multiple-choice moodle quiz)
- Also, course changes every year and some questions in old papers might cover topics that you wouldn't be asked about this year
  - e.g. question 1(d) is a bit out of scope
- This year's exam will be online and so won't ask questions that are completely based on memory such as defining terms
- These answers are bullet-point style for a marking scheme. You should write more, in proper sentences etc.

Question 1

- 1. This question is about visual usability and affordance.
  - (a) Describe what is meant by internal and external consistency. Give an example of both kinds of consistency in the context of a word processing application. **[4]**
  - (b) Give examples of two visual features that influence the perception of hierarchy. **[2]**
  - (c) Define the concepts of proximity and scale. Provide an example of a common mistake that decreases visible usability for each concept. **[4]**
  - ** (d) Describe the two influences that led to the development of an embodied approach to interaction design. **[4]**
  - (e) Define affordance and give an example of affordance in the context of a cash machine interface. **[4]**
  - (f) Define physical constraints and give an example in the context of a cash machine interface. **[2]**

Question 2

- 2. This question is about human factors.
  - (a) Provide a diagram for the model of human factors during interaction. **[8]**
  - (b) Describe the four bands of human action, giving an example for each. **[8]**
  - (c) Describe two aspects of human hearing, giving an example for how each is quantified or perceived. **[4]**

Question 3

- This question is about experimental design and evaluation methods.
  - **Consider the following experimental design. The experiment is about evaluating the impact of latency on the usability of soft keyboards. Participants are recruited from students at a university. During the experiment, participants experience one of three keyboard designs; keyboard with no added latency, keyboard with 50 milliseconds added latency, or keyboard with 100 milliseconds added latency. Each participant is asked to type out the same set of sentences in the same order. The experimenter records typing speed and errors.**

Question 3

- (a) Identify the independent and dependent variables in this experimental design. **[2]**
- (b) Discuss a benefit and a detriment of the between-subjects design of this experiment. **[4]**
- (c) Propose an alternative design using a within-subjects design, including how you would order the conditions. **[4]**
- (d) Identify a potential issue with this experimental design. State what kind of issue it is and how you would correct this issue. **[2]**
- (e) Design a closed-ended question that could be used to gather additional information from participants during this study. **[4]**
- (f) Describe two key aspects of consent that are required for ethical evaluations. For each aspect, state why this is important for ethical practice. **[4]**

Thanks!