Human Evaluation & Ethics

Research Skills and Professional Issues

Dr Simon Harper

School of Computer Science
Why Are You Here?

• Testing is key to science and research
• Testing with humans is normally done really badly
• Understanding the problems takes time and experience
• Building good methodologies allows our systems to be tested better, and therefore BE better
• Good Ethics makes: good methodologies; strong science; better systems

In the next 4 hours you’ll find out why and how
Timetable

- 4h Face to Face + 6h Post Lecture Practical
- Introducing Human Evaluation
- Your Past Experiences (P)
- Methods for Evaluation
- BREAK
- Augmenting Your Past Experiences (P)
- Ethical Applications
- Re Building Your Past Work
  - Practical over Week
Human Evaluation

- Human Computer Interaction
- Evaluation is the final part of the HCI process
- Often in a Usability Lab
- Based on Well Understood Science
- Participant is...
- Experimenter is...
Refutability

• Assertion must have the possibility of being Falsified
• Does NOT mean it IS False
• Just that it can be REFUTED

• Egs
  ❌ The interface will render in about 10ms
  ✔ The interface renders in 10ms
1. Create a refutable hypothesis
   - ‘All Swans are White’

2. Empirical observations which are **aimed at Refuting the hypothesis**
   - Observe as many Swans as possible in all countries and continents

3. Hypothesis is **Supported** (not proved) until a counter case is found
   - Arriving in Australia we find a Black Swan!

4. Therefore, ‘All Swans are White’ is false.
Requirements Analysis

• Most HCI collection and evaluation protocols are based on software engineering techniques

• Elicitation
  ✴ Gather data pre-implementation

• Analysis
  ✴ Analyse gathered data to create a systems design

• Systems approaches are often inflexible

• Where is the SE User Testing?
  ✴ Models created in system design are tested - often without further human intervention
**Significance of HCI**

- Humans are First-Class Citizens
  - Not after-thoughts
- System Conforms to User
  - Humans should not need to conform to the system
- Humans should be involved at All Stages
  - Not Just the Requirements Analysis
- SE Modelling techniques are sometimes not best suited to HCI - Too Engineering Focused
  - Methods from the Human Sciences often are suited
- Build flexible bespoke methods conforming to the scientific method
1. You have 45 minutes to:
   - Think back to your last implementation/project for a system or interface to be used by a human.
   - Think about how you either, tested it, or came up with the user requirements.
   - Try not to embellish - at this point just tell-it as it happened.

2. Create a 1 to 2 page document (Bullet Points):
   - Giving a brief overview of the project and its aims.
   - What data collection / testing techniques you used?
   - What kinds of things did you measure?
   - How did you analyse the data?
   - Did you make any changes based on the data?
   - What would you do differently?

• I’ll be picking on some of you for your analysis, next.
✔ What were the important things you DID RIGHT?

✗ What were the important things you DID WRONG?
One View of the HCI Process

1. **Understand**ing Users
2. Creating a Design
3. **Test**ing a Design
4. Building a Prototype
5. **Test**ing the Prototype
6. Building the Interface or System
7. **Evaluate**ting the Interface or System
8. Changing the Interface or System

Back to 1.

**Really About**: Testing and evaluating our understanding of users, captured within the design, prototype, or implementation.
• **Research Methodologies**

• **Methods for Pre Design Data Collection**
  - In the Field - Qualitative
  - In Between - Quantitative

• **Methods for Post Design Evaluation**
  - In Between - Quantitative
  - In the Laboratory - Experimental
  - Hybrid - Experimental with aspects of Qualitative & Quantitative
In the Field

- Participant Observation
- Interviewing or Conversation with a Purpose
- Focus Groups
- Archival Methods
- Unobtrusive Methods

- **Key-point: Field Notes**
  - A/V or Text
  - Immediate
  - Minor Delay
  - Day-Delay
  - Impressionistic
In Between

- Survey
- Questionnaires
- Walkthrough

- Key-point: Drawing the Sample
  - Random (p)
  - Systematic (p)
  - Stratified (p)
  - Multistage (p)
  - Quota (np)
In Between

- Think-Aloud
- Co-Operative Evaluation
- Participatory Design

+ Combinations of All

- Key-point: Involve People
  - Talk Out Loud
  - Participation
  - ‘Extreme Evaluation’
In the Laboratory

- Single Group, Post Test
- Single Group Pre & Post Test
- Natural Control Group Pre & Post Test
- Random Control Group Pre & Post Test
- Within Subjects
Key-point: Sample

- Population Definition
- Population
  - Census
- Sample Frame
  - Sample
- Probabilistic
  - Random
  - Systematic
  - Stratified
  - Multistage
- Non-Probabilistic
  - Quota
    - Demographics (Age, Gender, etc to match Popn Defn)
Key-point: Variables

- Subject, Behavioural, Stimulus, and Response
- Independent Variables
  - Thing Under Investigation (being manipulated or changed)
- Dependent Variables
  - Things that Depend (the observed result)
- Confounding Variables
  - Things that can lead to incorrect responses.
- Constants
  - Variables which may affect the others but which can be maintained as constant
Key-point: Measurement

- **Nominal**
  - Denotes identity
  - eg – Rod, Jane, Freddy

- **Ordinal**
  - Denotes identity and magnitude
  - eg – Enum {1, 20, 80}

- **Interval**
  - Denotes identity, magnitude, and has equal intervals
  - eg – Sequence

- **Ratio / Score**
  - Denotes identity, magnitude, has equal intervals, and a true zero point
  - eg – Integers
Key-point: Validity

- **Internal Validity**
  - Descriptive
  - Internal Consistency
  - Measuring Outliers

- **External Validity**
  - Inferential
  - Show your experiment can be generalized to a population
  - Measuring goodness of fit to that population
Instruments

• **Performance**
  * Time required by the user to complete a task
  * Time spent navigating the interface
  * Number of incorrect choices or errors created
  * Number of jobs completed, either correctly or incorrectly
  * Number of observations of user frustration
  * Frequency of interface components or behaviour that is never used.

• **Eye Tracking**

• **Facial Expressions**

• **Biofeedback**
  * GSR
  * Temperature
  * Blood Pressure
  * Heart Rate
Data Analysis (Experimental)

- **Inferential (Not Covered Any Further)**
  - Parametric
  - Non-Parametric

- **Descriptive**
  - Normal Distribution
Descriptive #1

• Central Tendency
  * Mode - Most frequent
  * Median - Middle Score
  * Mean - Arithmetic Average

• Variability
  * Skew - negative / positive
  * Kurtosis - mesokurtic / leptokurtic / platykurtic
  * Range
  * Standard Deviation
  * Average Deviation
  * Variance (Average Squared Distance from the Mean)

• Standard (Z-) Score
  * Describes how a single participant score relative to the rest of the participants (in relation to the mean)
• Relationships
  ※ Correlation
    - Correlation does not imply causation
    - Pearson Product Moment – for Score Data
      (Describes Linear Relationship)
    - Spearman Rank-Order – for Ordered Data
  ※ Regression
    - Predicts the value of one variable from the value of another
    - Scatter Plots
    - Line is called the Regression Line
    - Computationally easy on SPSS / PSPP
1. You have 1 hour to:

- Take the result of Practical #1; addressed in a laboratory setting
- Augment it with actual testable hypotheses
- Augment your methodology with descriptions of how you will make it conform to the scientific method
- What will be the variables tested etc.
- Demonstrate how you would support your hypotheses

- I’ll be picking on some of you for your analysis, next.
Back in 20 minutes - PROMPT
Practical #2 RECAP

1. You have 40 minutes to:
   - Take the result of Practical #1; addressed in a laboratory setting
   - Augment it with actual testable hypotheses
   - Augment your methodology with descriptions of how you will make it conform to the scientific method
   - What will be the variables tested etc.
   - Demonstrate how you would support your hypotheses

   - I’ll be picking on some of you for your analysis, next.
✔ What are the important things to CONSIDER?

✖ What did you MISS in your Original Methodology?
Ethics & Users

• Morals
• Ethics
• Standards Bodies

• Institutional Review Board
• Risk Assessment
• Informed Consent
• Privacy
Why Care?

- Abuse
- Subjects and Participants
- Keeping Us Honest
• Competency
  ✴ Keep up to date, know your limitations, ask for advice

• Integrity
  ✴ Have no axe to grind, or desired outcome;

• Scientific
  ✴ Follow the Scientific Method
  ✴ benefit from the results of that research
Principles #2

• **Respect**
  * Assess your participants autonomy and capability of self-determination, treat participants as equals, ensure their welfare

• **Benefits**
  * Maximising benefits and minimising possible harms according to you best judgement, seek advice from your organisations ethics committee

• **Justice**
  * Research should be undertaken with participants who will benefit from the results of that research

• **Trust**
  * Maintain trust, anonymity, confidentiality and privacy, ensure participants fully understand their roles and responsibilities and those of the experimenter

• **Responsibility**
  * You have a duty of care, not only to your participants, but also to the community from which they are drawn, and your own community of practice
Post Lecture Practical

• Read the hand out, and go over the Self Assessment Questions
• Divide into groups of 5
• From your 5 pick the most appropriate Practical #2 result
• Adapt the methodology to take into account the ethical procedures in the Handout
• Paying particular attention to Section 5.4
• Respond to each of the itemised points (there may be some duplication)
• Prepare to discuss this at length.