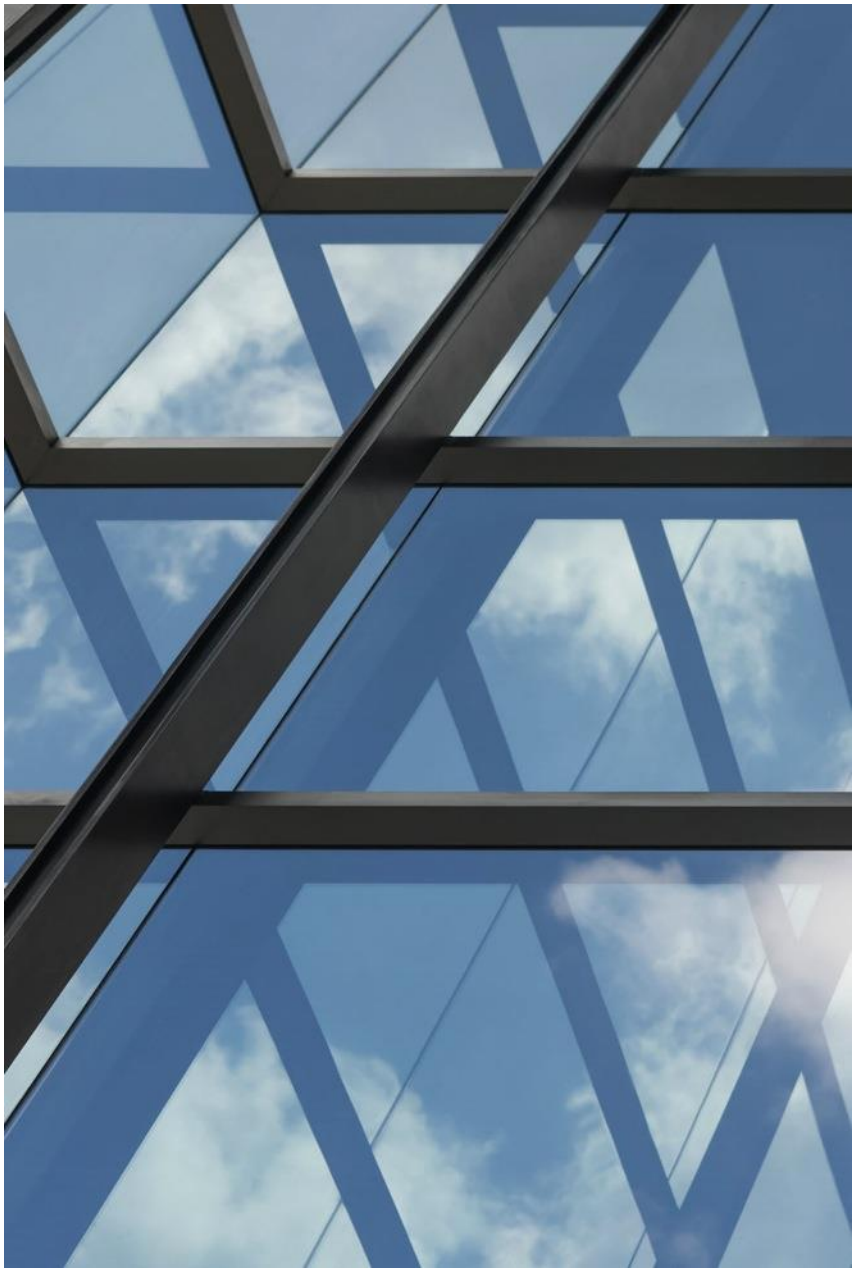
INFO3315 Human-Computer Interaction

Lecture 8 – Evaluation II

Zhanna Sarsenbayeva

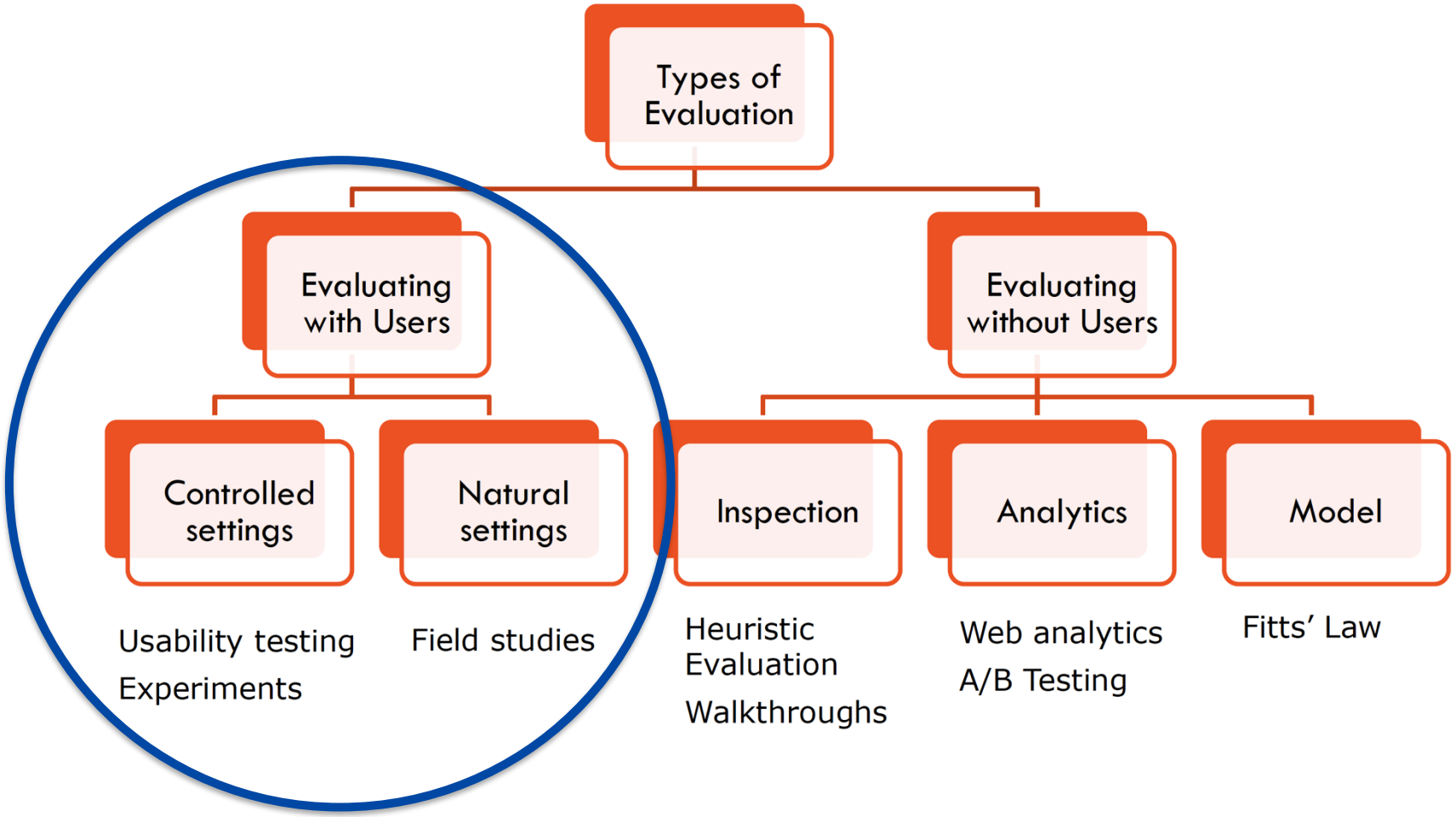


THE UNIVERSITY OF SYDNEY

Recap

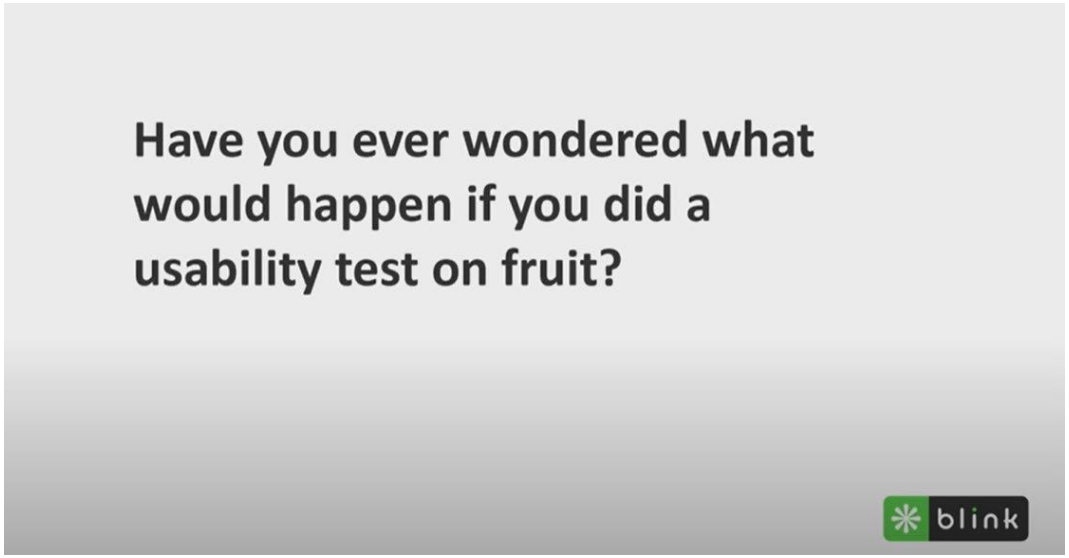
- Physical prototyping
- Evaluations: What? Where? When?
- Naturalistic vs. Lab environments
- Types of evaluation:
- Inspections – Heuristics/ Walkthroughs
- Analytics – Web, A/B Testing
- Models
- Cognitive Walkthrough example

Types of evaluation



Usability testing

- Involves recording performance of typical users doing typical tasks
- Users are observed and timed
- Data is recorded on video, and key presses are logged
- User satisfaction is evaluated using questionnaires and interviews



https://www.youtube.com/watch?v=3Qg80qTfzgU

Practical questions

- What should my team be doing?
  - Team roles during testing
- What should the participant be doing?
  - Task and task description
- Who should I recruit to be my participant?
  - Recruiting participants
- How many participants do I need?
  - Number of participants
- How do I prepare for a test session?
  - Planning the session
  - Paperwork – important (informed consent) and others

What should my team be doing?

- All members should be encouraged to participate in the evaluation.
- Roles during usability test:
  - Facilitator
  - Prototype executor
  - Quantitative data collector
  - Qualitative data collector
  - Supporting Actors

Facilitator

- Person in the lab together with participant
- Responsibilities:
  - Plan and execute session
  - Set up lab for session
  - Responsible for putting participant at ease during session
  - Must have people skills



https://youtu.be/3Qg80qTfzgU

Prototype executor

- Person to “execute” the prototype and move it through its paces as users interact
- Only if you are using a low-fidelity prototype
- Must have thorough technical knowledge of how design works
- Poker face and should not speak a single word during the session.

Quantitative data collectors

- If you plan to collect quantitative data, you will need people to do it
  - Time to complete task
  - Number and type of errors per task
  - Number of errors per unit of time
  - Number of navigations to online help or manuals
- Record into spreadsheet directly (lo-fi prototype)
- Tools – stopwatch and counter
- Can be programmed (hi-fi prototype)

Table 10.3 Common usability metrics

Usability objective	Effectiveness measures	Efficiency measures	Satisfaction measures
Overall usability	• Percentage of tasks successfully completed • Percentage of users successfully completing tasks	• Time to complete a task • Time spent on non-productive actions	• Rating scale for satisfaction • Frequency of use if this is voluntary (after system is implemented)
Meets needs of trained or experienced users	• Percentage of advanced tasks completed • Percentage of relevant functions used	• Time taken to complete tasks relative to minimum realistic time	• Rating scale for satisfaction with advanced features
Meets needs for walk up and use	• Percentage of tasks completed successfully at first attempt	• Time taken on first attempt to complete task • Time spent on help functions	• Rate of voluntary use (after system is implemented)
Meets needs for infrequent or intermittent use	• Percentage of tasks completed successfully after a specified period of non-use	• Time spent re-learning functions • Number of persistent errors	• Frequency of reuse (after system is implemented)
Learnability	• Number of functions learned • Percentage of users who manage to learn to a pre-specified criterion	• Time spent on help functions • Time to learn to criterion	• Rating scale for ease of learning

Source: ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs), extract of Table B.2

Qualitative data collectors

- Remaining team member should be ready to collect qualitative data.
- Example data to be recorded:
  - Observation notes
  - Critical incidents
  - Think aloud comments

Supporting actors

- If part of the setting or task requires participant to interact with someone.
- Manage the props needed in the evaluation (other than the prototype execution).
- Example: Call client on the telephone.

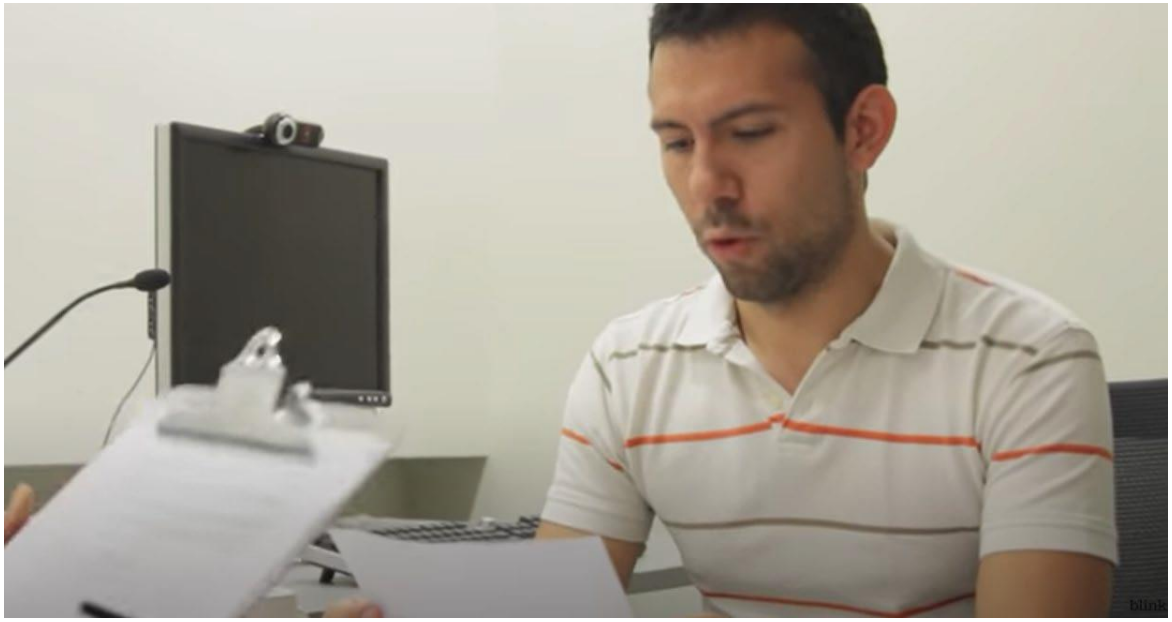
Tasks during session

- Usability tests are usually task-based.
- Representative, frequent and critical tasks that apply to the key work role and user class represented by each participant.
- Prepare corresponding task description and UX target metrics to guide data collection and compare observed results.
- Test conditions same for every participant

Task description

- Task description – what to do, no hints about how to do

- Example from the video:
  - *Show me how you eat this*



https://youtu.be/3Qg80qTfzgU

Recruiting participants

- Find representative users (usually outside your team and outside project organization)
- Recruitment methods and screening:
  - People around you – spouses, children, friends
  - Post ads in public spaces
  - Announcements at meetings of user groups and professional organizations if the group matches your user class needs
  - Temporary employment agencies

Number of participants

- Number is a practical issue
- Depends on:
  - Schedule for testing
  - Availability of participants
  - Costs of running tests
- Famous rule of thumb – three to five participants
  - Origin: Five participant evaluators is sufficient to find approximately 80% of UX problems in a system
  - Does not always apply to all situations
- Typically 5-10 participants
- OR: Test until no new insights are gained

Planning the session

- If it is in the lab, configure the lab to your needs.
  - Computer/Device
  - Placement of participant, facilitator and executor
- Set up hardware e.g. eye-trackers, timers, counters etc.
- Determine length of session for one participant
- Prepare necessary paperwork
  - Informed consent (important)
  - Other data collection forms

Length of session

- Typical length: 30 mins to 120 mins
- Strategies to manage long sessions:
  - Warn participants in advance
  - Schedule breaks between task – exercise, toilet break, refreshments
  - Prepare food and water in advance to keep participant at ease

Informed consent

- Formal and signed permission given to UX professional by participant to use data gathered within stipulated limits.
- Preparation for informed consent begins with institutional review board (IRB) / Ethics Approval Committee
- Evaluator/Project manager to prepare application.

USYD Ethics Application

- Participant Information Statement (PIS)
- Participant Consent Forms (PCF)
- Advertisements, letters and emails seeking participants
- Interview or focus group questions/themes
- Letters of support or permission from organisations assisting in the research in any way
- External researcher declarations (for researchers not affiliated with the University)

PIS and PCF templates

Participant Information Statement Templates	Participant information statement (General, DOC 66KB) Participant information statement (Parent/carer, DOC 140KB) Participant information statement (Person responsible, DOC 143KB) Participant information statement (Easy English/child, DOC 487KB) Participant information statement (Interview/oral history, DOC 53KB) Participant information statement (Observation/ ethnography/ anthropology, DOC 55KB) Participant information statement (clinical trial, DOC 62KB)
Participant Consent form samples	Participant consent form (General, DOC 55KB) Participant consent form (Parent/carer, DOC 56KB) Participant consent form (Person responsible, DOC 57KB) Participant consent form (Easy English/child, DOC 49KB) Participant consent form (Interview/oral history, DOC 53KB) Participant consent form (Observation/ethnography/anthropology, DOC 57KB)

Informed consent

- Two parts:
  - Participant information statement (PIS)
  - Participant consent form (PCF)
- Participant must read both PIS and PCF before session
- Allow participant/guardians to ask questions.
- Participant/guardian must sign PCF before session.
- Prepare two copies for the session:
  - One for participant to keep
  - One for submission

Other data collection forms

- Non-disclosure Agreements (NDAs)
  - If required by developer or customer organizations to protect intellectual property(IP) contained in design.
  - Must be included during signing of PCF.
- Questionnaires
  - If your evaluation plan includes administration of one.
  - Pre-questionnaire – before study. Post-questionnaire – after study.
  - Example questionnaires for usability:
    - System Usability Scale (SUS)
    - Usefulness, Satisfaction and Ease of Use (USE) questionnaire

On the big day

- **Before Session**
- Invite participant into the lab
- Offer refreshments and paperwork (PIS, PCF and questionnaire)
- Explain details of study and check for questions
- Participant to complete requested forms.
- [Optional] Interview participant to check responses on questionnaire

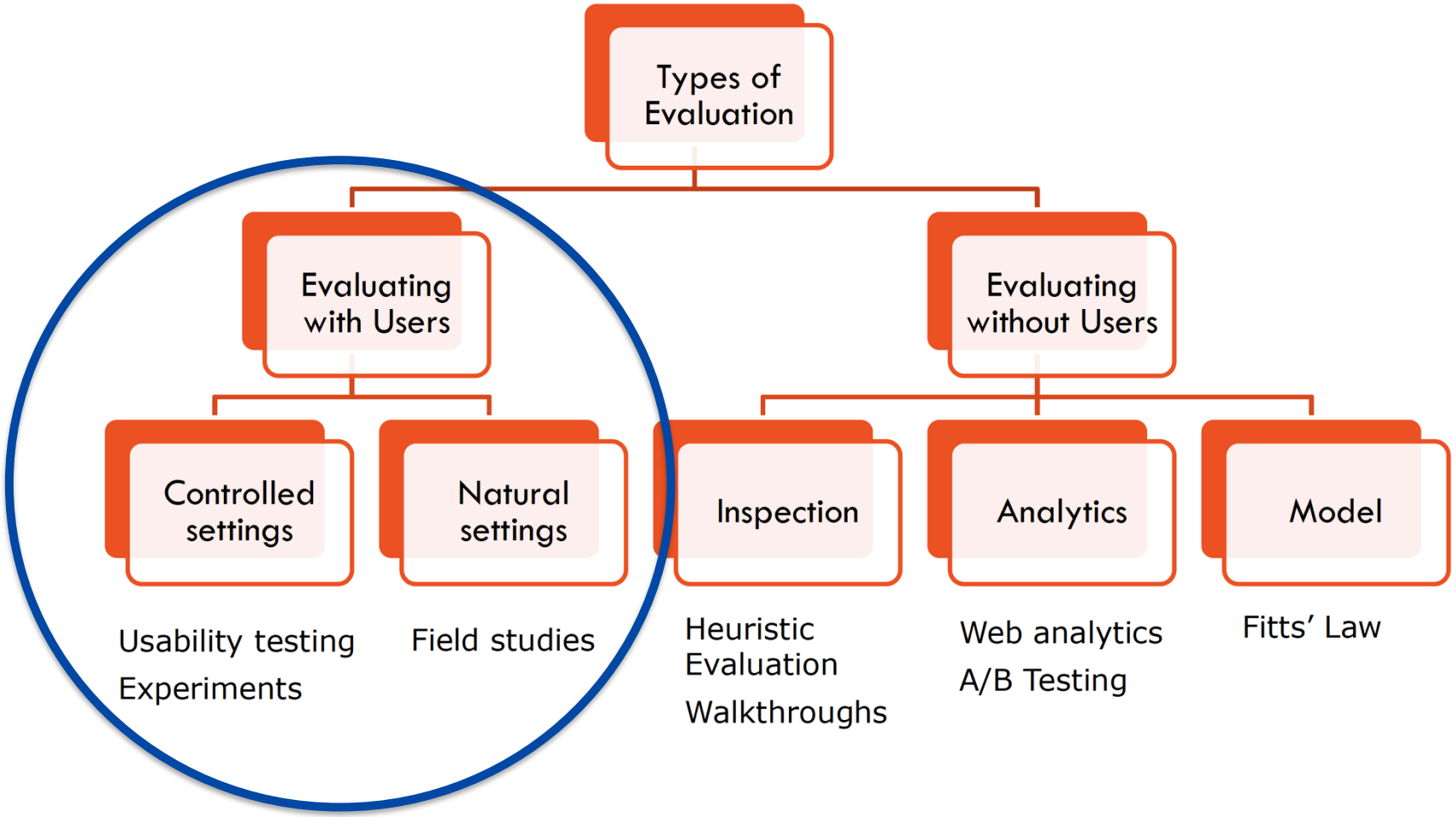
On the big day

- **During Session**
- Hand out task (one at a time)
- Encourage participant to think-aloud
- Describe his actions and why
- Stop test if participant is in distress
- To help or not to help participant during the task:
- Depends on purpose of test
- Guide users if questions are asked

On the big day

- **End of Session**
- Post-session probes if you have any.
- Debrief participant -answer remaining questions that the participant has and what you will do next.
- Thank the participant and give token of appreciation for their time.
- Prepare for the next participant.

Experiments



Usability testing and Experiments

- Usability testing is applied experimentation
- Developers check that the system is usable by the intended user population by collecting data about participants' performance on prescribed tasks
- Experiments test hypotheses to discover new knowledge by investigating the relationship between two or more variables

Usability testing vs. Research

- Usability testing
- Improve products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

- Experiments for Research
- Discover knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strictly controlled conditions
- Experimental design
- Scientific report to scientific community

Week 8 – In-class quiz

- 5 mins
- 4 questions
- Need to score 1 point

Experiments

- Test hypothesis
- Predict the relationship between two or more variables
- Independent variable is manipulated by the researcher
- Dependent variable influenced by the independent variable
- Typical experimental designs have one or two independent variables
- Typical experimental designs have one or two independent variables
- Validated statistically and replicable

Experimental designs in HCI

- Humans learn from experience
- When dealing with human subjects, need to determine which participants to involve for which conditions in an experiment
- Experience in participating in one condition will affect performance of those participants if asked to participate in another condition

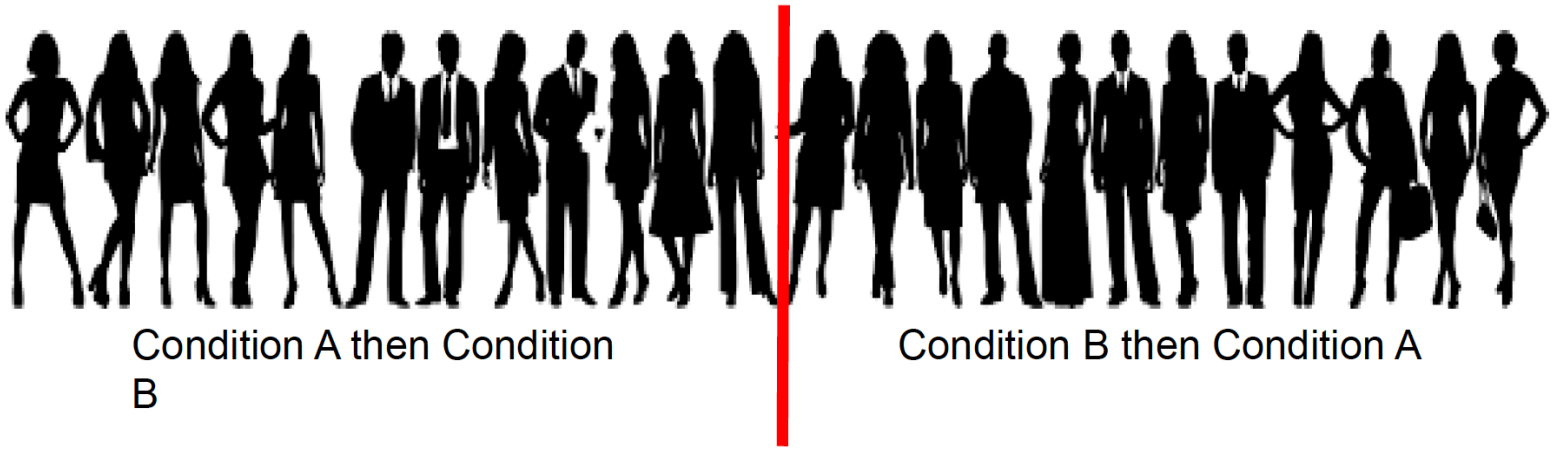
Experimental designs

- Different-participants design (between-subjects design)
- Single group of participants is allocated randomly to each of the experimental conditions so that different participants perform in different conditions.
- Need many participants to minimize individual differences – differences in experience and expertise. Perform pre-testing to identify participants that differ strongly.



Experimental designs

- Same-participants design (within-subjects design)
- All participants perform in all conditions
- Must perform counter-balancing to reduce effects of ordering.
- Effects of ordering – learning from previous task affects performance of subsequent task.



Experimental designs

- Matched-participant design (pair-wise design) - Participants are matched in pairs based on certain user characteristics such as expertise and gender.
- Each pair is then randomly allocated to each experimental condition.
- Does not consider important variables that may influence the results

How to: Experiments

1. Determine goals, explore the questions then formulate hypothesis
2. Design experiment, define experimental variables
3. Choose subjects
4. Run pilot experiment
5. Iteratively improve experiment design
6. Run experiment
7. Interpret results to accept or reject hypothesis

Experiments

Effect of Distinct Ambient Noise Types on Mobile Interaction

ZHANNA SARSENBAYEVA, The University of Melbourne, Australia
NIELS VAN BERKEL, The University of Melbourne, Australia
EDUARDO VELLOSO, The University of Melbourne, Australia
VASSILIS KOSTAKOS, The University of Melbourne, Australia
JORGE GONCALVES, The University of Melbourne, Australia

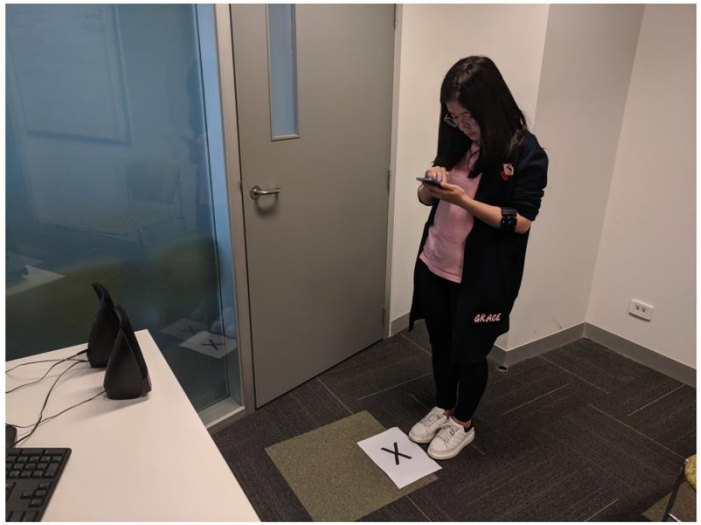
The adverse effect of ambient noise on humans has been extensively studied in fields like cognitive science, indicating a significant impact on cognitive performance, behaviour, and emotional state. Surprisingly, the effect of ambient noise has not been studied in the context of mobile interaction. As smartphones are ubiquitous by design, smartphone users are exposed to a wide variety of ambient noises while interacting with their devices. In this paper, we present a structured analysis of the effect of six distinct ambient noise types on typical smartphone usage tasks. The evaluated ambient noise types include variants of music, urban noise and speech. We analyse task completion time and errors, and find that different ambient noises affect users differently. For example, while speech and urban noise slow down text entry, being exposed to music reduces completion time in target acquisition tasks. Our study contributes to the growing research area on situational impairments, and we compare our results to previous work on the effect of cold-induced situational impairments. Our results can be used to support smartphone users through adaptive interfaces which respond to the ongoing context of the user.

CCS Concepts: • **Human-centered computing** → Empirical studies in HCI; • **Human-centered computing** → Ubiquitous and mobile computing; • **Human-centered computing** → Smartphones;

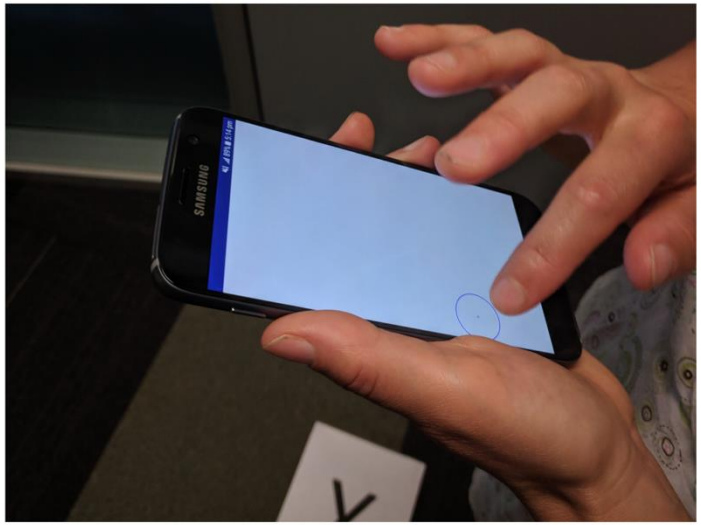
https://doi.org/10.1145/3214285

Experiment – Basis

- **Aim of the experiment:**
- To understand if different types of noise influence the way we interact with smartphones



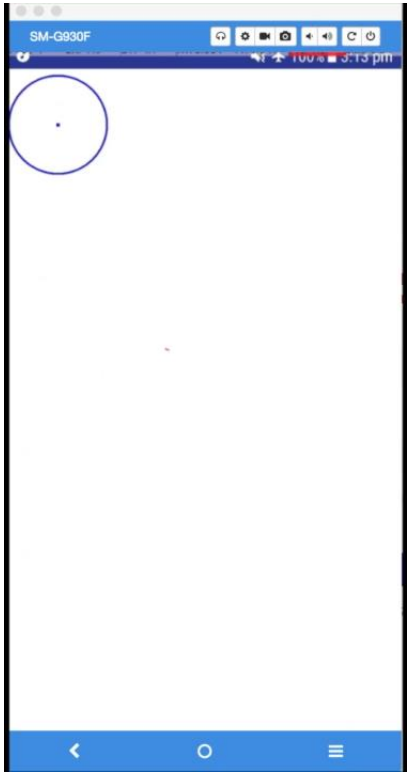
(a) Laboratory setting of the experiment

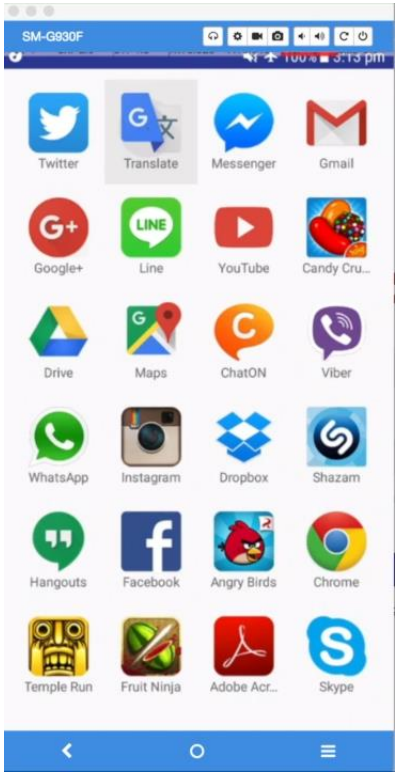


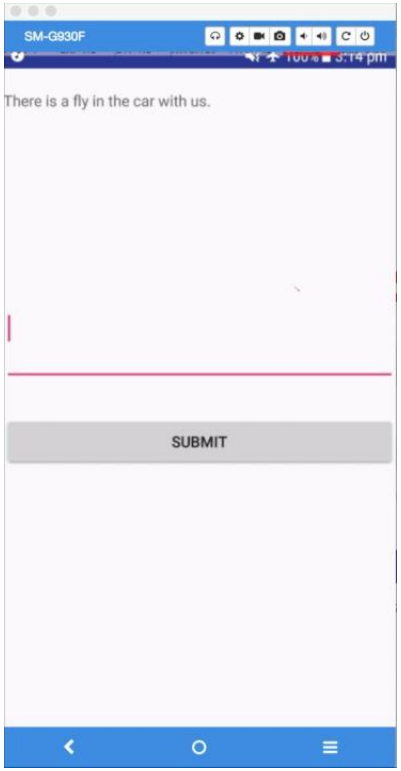
(b) Smartphone holding posture

Experiment – Basis

- Hypothesis:
  - Different types of ambient noise affect the interaction with mobile devices during three typical activities: target acquisition, visual search, and text entry







Experiment – Basis

- Controlled lab study:
  - 24 participants (12F, 12M)
  - Duration – 40 min per participant

Experiment – Basis

- **Types of noise**
- Music (classical): Fast tempo, slow tempo
- Urban noise: Indoor, Outdoor
- Speech: Meaningful (English), Meaningless (Kazakh)
- Silence: Acted as a control condition

- All types of noise were played to participants at 60dB

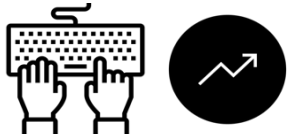
Experiment – Basis

- **Variables:**
- Independent variables: types of noise, participant gender, age
- Dependent variables:
  - Target acquisition task: task completion time, offset size of the touch
  - Visual search task: visual search time, memorization time, error rate
  - Text entry task: time per character entry, error rate

Experiment – Basis

- Procedure
- Greet the participant
- Brief the participant about the study, get signed consent
- Train the participant so they become familiar with the smartphone tasks
- Present the conditions to participants in a counter-balanced order (within subject design)
- Order of tasks randomised to avoid sequence effect
- Conduct a short semi-structured interview
- Debrief and reward the participant

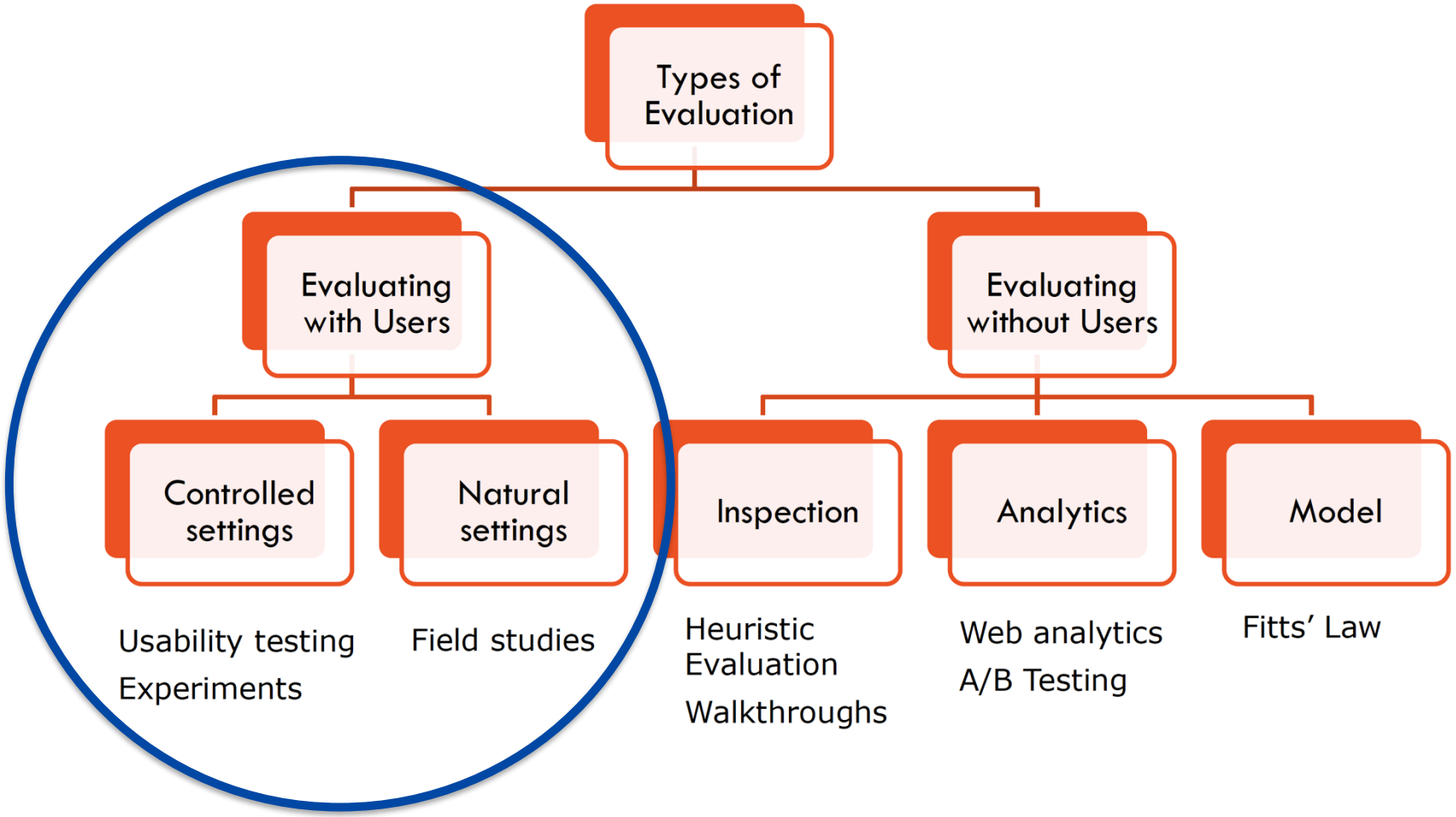
Our results

- Target Acquisition Time (Music: Fast and Slow) 
- Offset size (Music: Fast and Slow) 
- Time to Memorise Icons (Urban Indoor) 
- Errors in Visual Search (Urban Outdoor) 
- Time per Char Entry (Urban Outdoor, Meaningful Speech) 

To summarise:

- Hypothesis examines relationship (effect) between independent variable on dependent variables
- Independent variable is what the researcher manipulates e.g. test conditions
- Dependent variable is the change that is caused by the manipulation e.g. performance
- Experimental design –test hypothesis and set up the conditions to keep other variables constant to prevent them from influencing findings

Field studies



Field studies

- Field studies are done in natural settings
- "In the wild" is a term for prototypes being used freely in natural settings
- Seek to understand what users do naturally and how technology impacts them
- Field studies are used in product design to:
  - Identify opportunities for new technology
  - Determine design requirements
  - Decide how best to introduce new technology
  - Evaluate technology in use

Field studies

- Range from a few minutes to longitudinal studies (few years).
- Data collection that is unobtrusive but informative
  - Self-reports of problems encountered when occur
  - Interval logging triggered by smartphone notifications
  - Logging software – monitor frequency/patterns of daily activities

Conundrums with field studies

- Informing participants that they are studied
  - Knowledge of study would make people conscious about how they behave
  - Without informing people of their participation, how do you get their consent for participation?
- Privacy of participants
  - Studies done in people's homes will always be intrusive
  - Agreement between participant and researcher the activities that can or cannot be recorded
- What to do with the prototype?
  - Event of breakdown
  - Security arrangements if deployed in public spaces

Field studies

Does Smartphone Use Drive our Emotions or vice versa? A Causal Analysis

Zhanna Sarsenbayeva[1], Gabriele Marini[1], Niels van Berkel[2], Chu Luo[1], Weiwei Jiang[1], Kangning Yang[1], Greg Wadley[1], Tilman Dingler[1], Vassilis Kostakos[1], Jorge Goncalves[1]
[1] The University of Melbourne, Melbourne, Australia, firstname.lastname@unimelb.edu.au
[2] Aalborg University, Aalborg, Denmark, nielsvanberkel@cs.aau.dk

ABSTRACT
In this paper, we demonstrate the existence of a bidirectional causal relationship between smartphone application use and user emotions. In a two-week long in-the-wild study with 30 participants we captured 502,851 instances of smartphone application use in tandem with corresponding emotional data from facial expressions. Our analysis shows that while in most cases application use drives user emotions, multiple application categories exist for which the causal effect is in the opposite direction. Our findings shed light on the relationship between smartphone use and emotional states. We furthermore discuss the opportunities for research and practice that arise from our findings and their potential to support emotional well-being.

Author Keywords
Smartphones, emotions, mobile interaction, mobile application use, emotion detection, emotional well-being

INTRODUCTION
Emotions are integral to human behaviour and influence the and the adoption of technology [43], and asserting the importance of designing technology to create desirable emotional experiences [67]. Recently, researchers have studied this in more specific contexts. For example, the use of digital technologies as diverse as smartphones [61], videogames [20], social media [48], online videos [64], music streaming services [49] and virtual reality [94] are known to influence emotions.

Recent research has also examined the 'inverse' causal direction: how emotional states influence technology use. States of joy and anger have been shown to affect interaction, such as increase joystick movements, speed, and acceleration compared to states of sadness [4]. Even typing speed is affected by our emotional state: Khanna and Sasikumar [47] found that when experiencing negative emotional states users type slower compared to when experiencing positive states. Likewise, a user's emotional states influence interaction with mobile devices [13]. For example, Cairns *et al.* show in their study that users in a negative emotional state are more likely to make errors when completing a simple task of number entry on a tablet [13]. Moreover, Mottelson and Hornbæk [63] demonstrate that a user's emotional state has a behavioural effect on mobile interaction.

https://doi.org/10.1145/3313831.3376163

Field study – Basis

- **Aim of the experiment:**
- To understand if smartphone usage drives people emotions or if it is vice versa



https://www.lifehack.org/427248/10-innovative-ways-of-using-your-smartphone

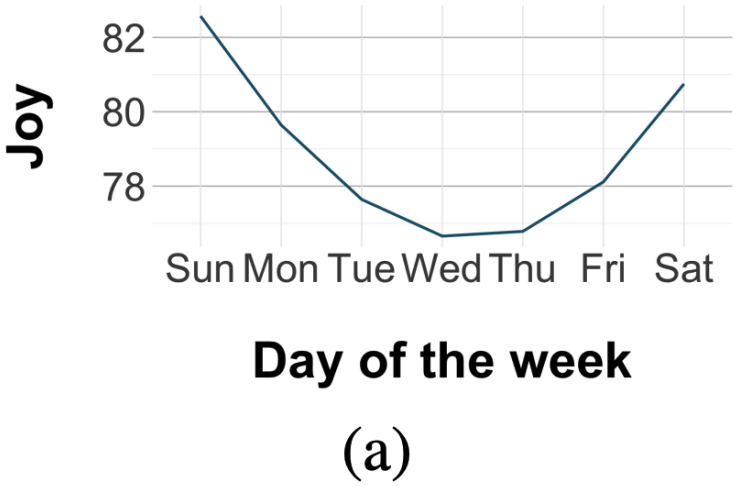
Field study – Basis

- Developed a standalone Android application
  - Tracks application usage
  - Tracks users facial expressions and logs recorded emotions (Affectiva API)
  - Collects self-reported emotions through questionnaires (6 times per day)
  - Runs as a service on the background (not so easy to do this after Android update in September 2019)
- 2 weeks study
- 30 participants (15F, 15M)

Field study – Basis

- **Variables:**
- Independent variables
  - Participant gender, day of the week, application use (duration, application type),
- Depended variables
  - Emotions (sadness, joy, fear, anger, surprise and disgust, valence), self-reported emotions (valence, arousal)

Validation of self-reports



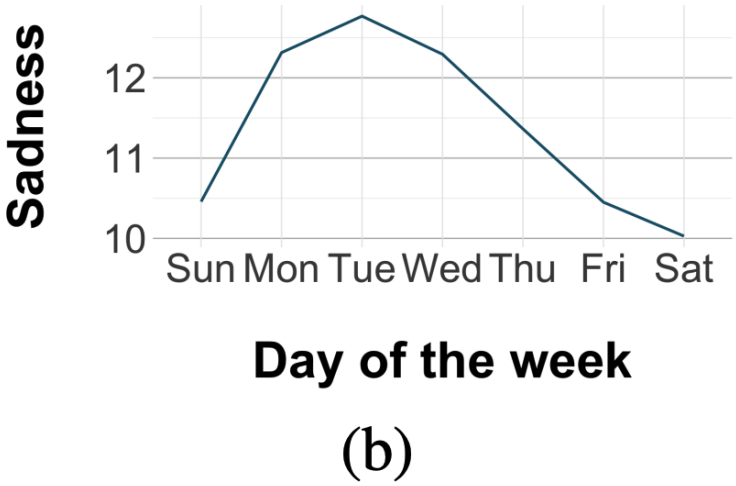


Figure 2: Daily fluctuations for joy and sadness throughout the week (smoothed conditional means)

ESM valence positively correlates with the mean Affectiva valence values ($r = 0.3$, $p < 0.01$)

Our results

- Convergent Cross Mapping (CCM) method (Sugihara et al., 2012)
- For some participants, the use of particular apps causes them to experience certain emotions
- However, for other participants, their emotions drive app use behaviour

Metric	Contempt	Disgust	Joy	Sadness	Surprise
Total Apps Usage Duration	0.091 (0.172)	0.047 (0.220)	0.138 (0.191)	0.086 (0.214)	0.144 (0.247)
Communication Apps Usage Duration	-0.118 (0.534)	0.131 (0.393)	0.124 (0.331)	0.122 (0.388)	0.177 (0.363)
Social Apps Usage Duration	0.221 (0.387)	0.093 (0.680)	0.129 (0.360)	0.168 (0.215)	-0.025 (0.469)
Work Apps Usage Duration	-0.060 (0.558)	0.064 (0.574)	-0.180 (0.316)	0.034 (0.547)	0.265 (0.315)
Entertainment Apps Usage Duration	0.149 (0.302)	-0.010 (0.616)	0.256 (0.471)	-0.050 (0.680)	-0.359 (0.379)

Table 2: Effect Sizes (and SD) for Causality between Application Usage Duration and Emotions. Blue: phone use drives emotions; Orange: emotion drives phone use.

To summarise

- Field studies are evaluation studies that are carried out in natural settings to discover how people interact with technology in the real world
- Sometimes the findings of a field study are unexpected, especially for in-the-wild studies that explore how novel technologies are used by participants in their own homes, places of work, or outside

Required reading

- Preece, J., Rogers, Y., & Sharp, H. (2019). Interaction design: beyond human-computer interaction (Fifth edition.). Wiley. Chapter 15