v1.0

Field Studies in Ubiquitous Computing

Tilman Dingler

Mobile Computing | 15-Oct-2019

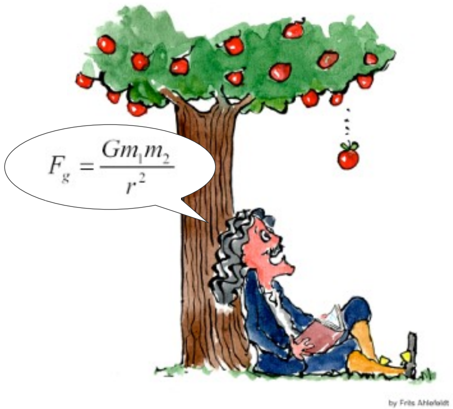
COMP90018 - Mobile Computing Systems Programming

X

Collecting Empirical Evidence

- How do the planets move?
- Who uses Android phones?
- How many internal emails do company workers receive on a daily basis?

Deriving Laws & Principles



- Why do planets move in specific orbits?
- Does income influence choice of smartphone?
- Does email communication improve productivity?

Knowledge Gain

1. Types of field studies

2. How to design and conduct a study in-the-wild

3. Making sense of study data

Learning Outcomes

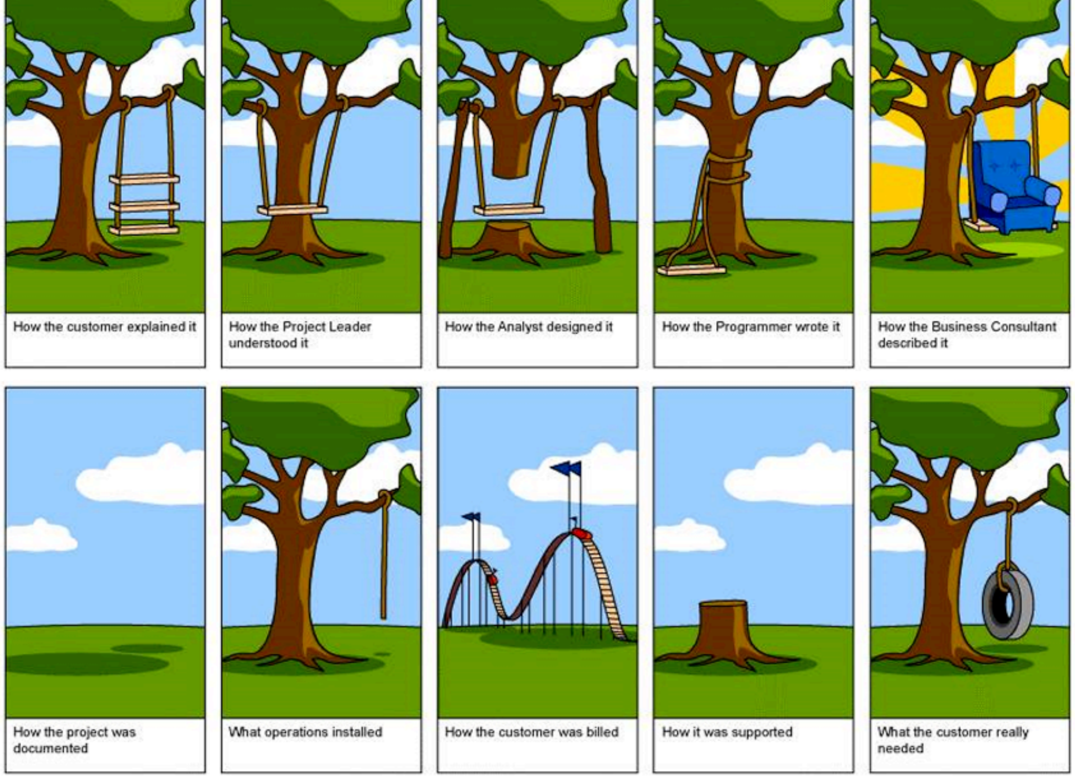
1. Formulate a research question
2. Come up with a hypothesis
3. Experiment design and plan
4. Conduct experiment, collect data
5. Data analysis and interpretation
6. Share results

Empirical Research

1. **A new system was developed**
   - Design goals reached?
   - Usability testing
   - Results not always generalisable

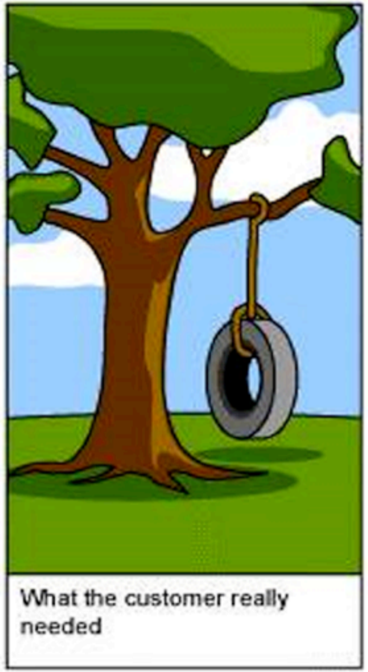
2. **A new system shall be developed:**
   - what are user constraints?
   - what are context constraints?
   - Results often generalizable

Identifying a Research Question

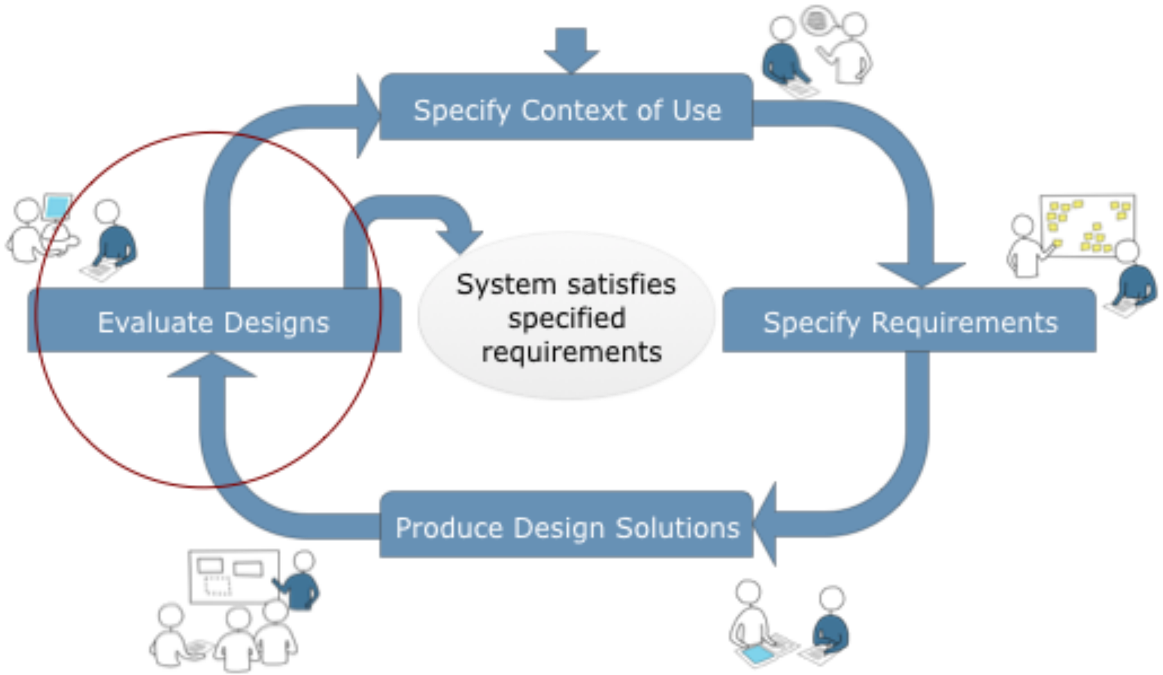


Developing a New System

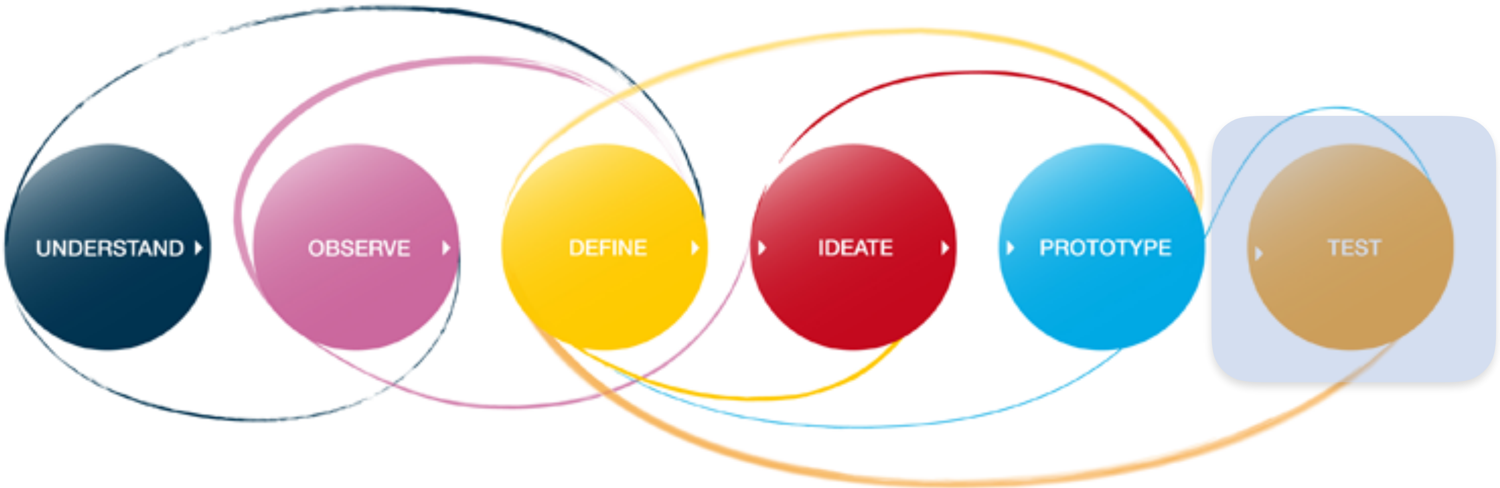
- **Lab studies**
  - Heuristical evaluation
  - Usability tests
  - High internal validity
- **Evaluation “in-the-wild”**
  - Proof-of-Concept
  - Experience using a prototype
  - High external validity



A New System was Developed



Human-Centred Design Process



Design Thinking

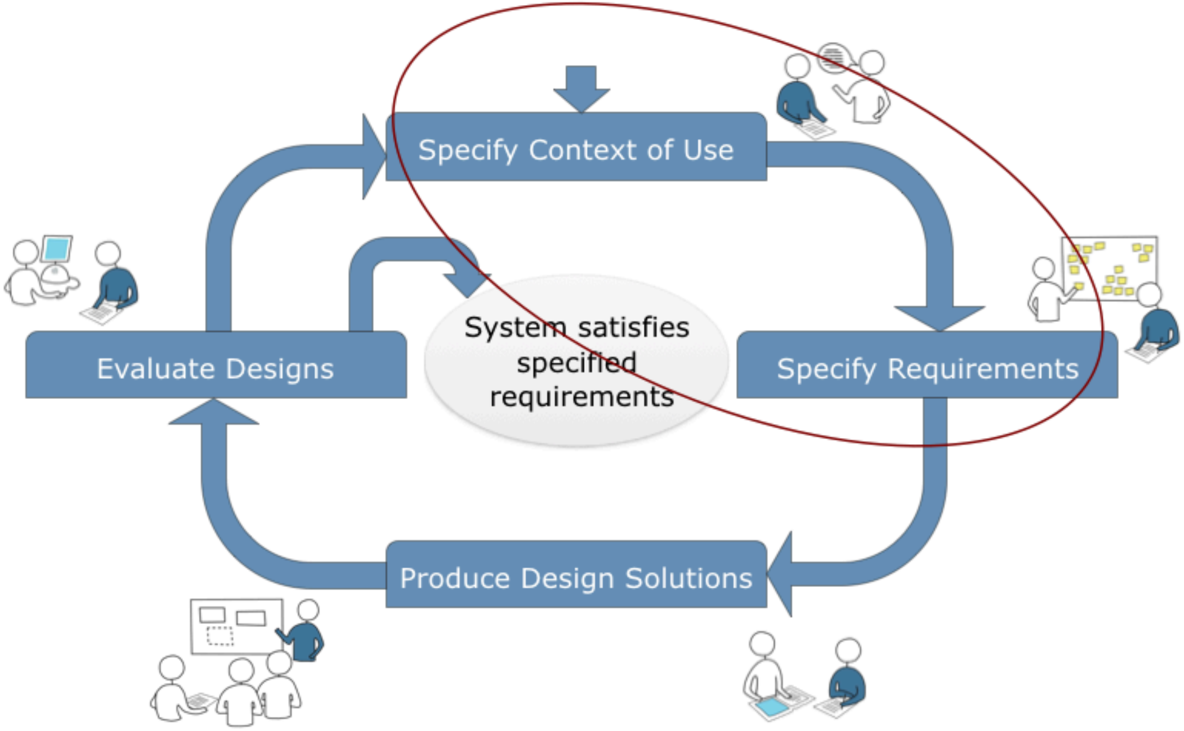
http://leschoolab.com/wp-content/uploads/2014/06/design_thinking_process.png

- Context of use and requirements analysis

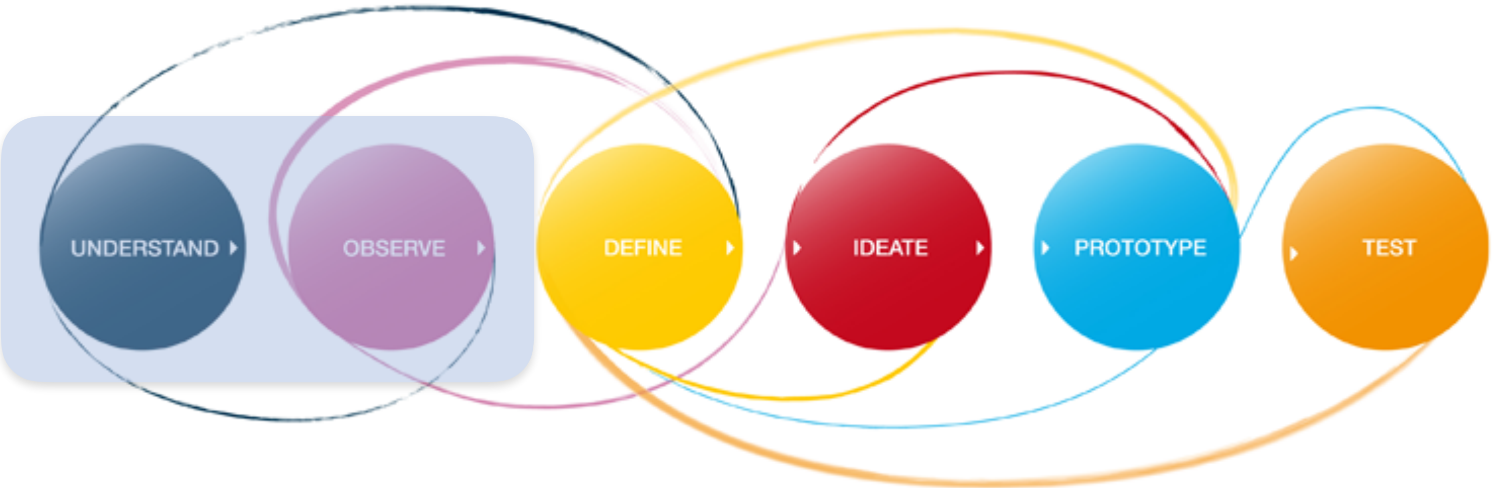
- What?
  - Understand as much as possible about users, tasks and contexts
  - Produce a set of requirements

- Why?
  - A high success factor
    - Increasing acceptance and satisfaction
    - Cost-reduction

Developing a New System



Human-Centred Design Process



Design Thinking

http://leschoolab.com/wp-content/uploads/2014/06/design_thinking_process.png

A user **study** conducted **outside** a research laboratory or controlled environment (i.e., "in the field").

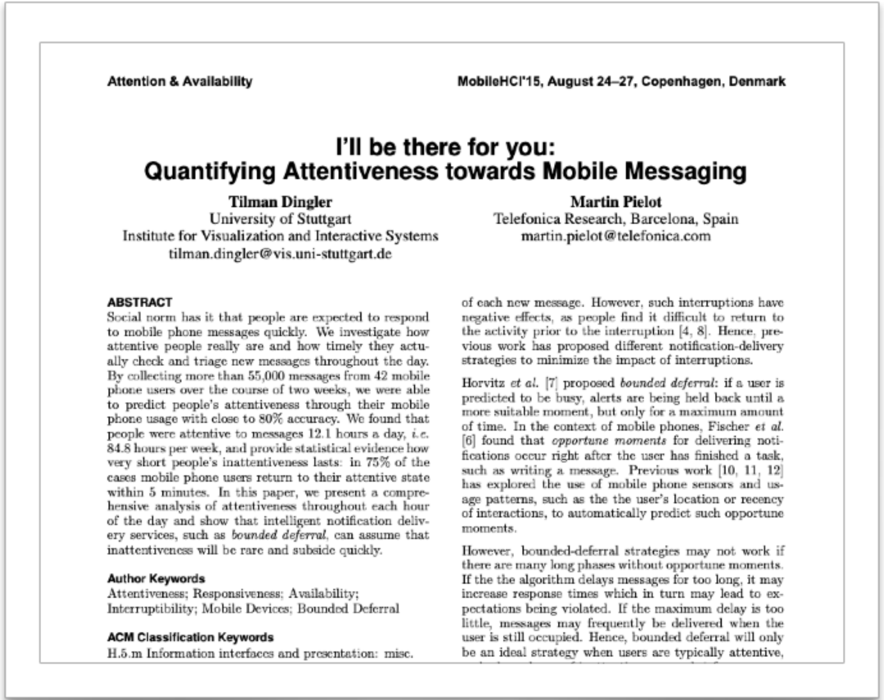
Field Study

- **Collect** abundant data about the use of technologies
- **Observe** unexpected challenges users may experience
- **Understand** how technology impacts uses’ lives
- Trade-off compared to a controlled lab study
  - Increased realism (external validity)
  - Loss of control over users’ experience

Field Study: Goals

1. Studies of **current behaviour**
   - What are people doing now?
2. **Proof-of-Concept** studies
   - Does my technology/app/system work in the real world?
3. **Experience** using a prototype
   - How does using my prototype change people's behaviour or allow them to do new things?

Field Study Types



I'll be there for you: Quantifying Attentiveness towards Mobile Messaging

Tilman Dingler
University of Stuttgart
Institute for Visualization and Interactive Systems
tilman.dingler@vis.uni-stuttgart.de

Martin Pielot
Telefonica Research, Barcelona, Spain
martin.pielot@telefonica.com

ABSTRACT
Social norm has it that people are expected to respond to mobile phone messages quickly. We investigate how attentive people really are and how timely they actually check and triage new messages throughout the day. By collecting more than 55,000 messages from 42 mobile phone users over the course of two weeks, we were able to predict people's attentiveness through their mobile phone usage with close to 80% accuracy. We found that people were attentive to messages 12.1 hours a day, *i.e.* 84.8 hours per week, and provide statistical evidence how very short people's inattentiveness lasts: in 75% of the cases mobile phone users return to their attentive state within 5 minutes. In this paper, we present a comprehensive analysis of attentiveness throughout each hour of the day and show that intelligent notification delivery services, such as *bounded deferral*, can assume that inattentiveness will be rare and subside quickly.

Author Keywords
Attentiveness; Responsiveness; Availability; Interruptibility; Mobile Devices; Bounded Deferral

ACM Classification Keywords
H.5.m Information interfaces and presentation: misc.

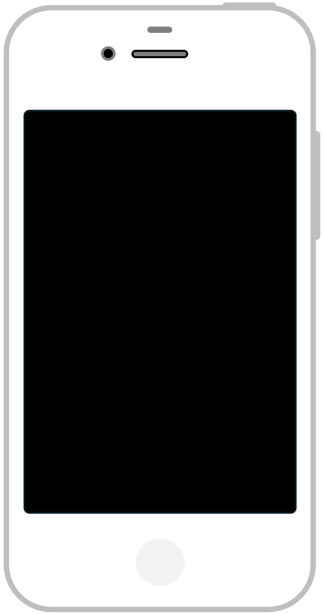
of each new message. However, such interruptions have negative effects, as people find it difficult to return to the activity prior to the interruption [4, 8]. Hence, previous work has proposed different notification-delivery strategies to minimize the impact of interruptions.

Horvitz *et al.* [7] proposed *bounded deferral*: if a user is predicted to be busy, alerts are being held back until a more suitable moment, but only for a maximum amount of time. In the context of mobile phones, Fischer *et al.* [6] found that *opportune moments* for delivering notifications occur right after the user has finished a task, such as writing a message. Previous work [10, 11, 12] has explored the use of mobile phone sensors and usage patterns, such as the the user's location or recency of interactions, to automatically predict such opportune moments.

However, bounded-deferral strategies may not work if there are many long phases without opportune moments. If the the algorithm delays messages for too long, it may increase response times which in turn may lead to expectations being violated. If the maximum delay is too little, messages may frequently be delivered when the user is still occupied. Hence, bounded deferral will only be an ideal strategy when users are typically attentive,

Studying User Behaviour

based on how people **use** their **phone** to derive **prediction** models **to estimate attentiveness** throughout the day



Investigating People‘s General Attentiveness

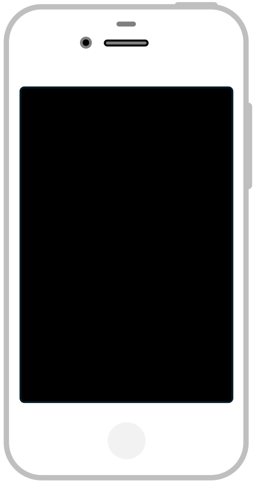
Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**.

- **Usage data** from **42 mobile** phone users over **2 weeks**
- Recruited through mailing lists of the university and a large IT organization
- **Demographics** voluntary
  - 45.1% male, 23.8% female, 31% unreported
- Mean reported age of 28.7 years (SD=5.9)
- Informed consent presented on first app launch
- 20 EUR voucher compensation

Participants

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI '15**.

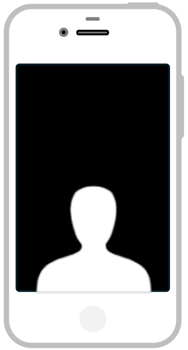
- Android App, available on *GooglePlay*
- **Background service** collecting
  1. Time of message arrival
  2. Time until attended to
  3. Contextual and phone-usage data
     - Screen on/off
     - Proximity sensor
     - Notifications
     - Ringer mode
     - App in foreground
- Trained a **machine-learning algorithm** to **infer message attendance**



Data Collection

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI '15**.

- Ground Truth: Messages received and attended to
- Machine Learning model based on 16 features
  - Screen status
  - Phone unlocks
  - Message data
  - Notification drawer
  - Time
  - Ringer Mode



Inferring Attentiveness from Phone-Usage Patterns

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**. http://dx.doi.org/10.1145/2785830.2785840

- Messages **attended** within a median time of **2.08min**
  - 25% within 12.0 seconds
  - 75% within 12.3 minutes
  - 95% within 80 minutes

Focus on Attentiveness

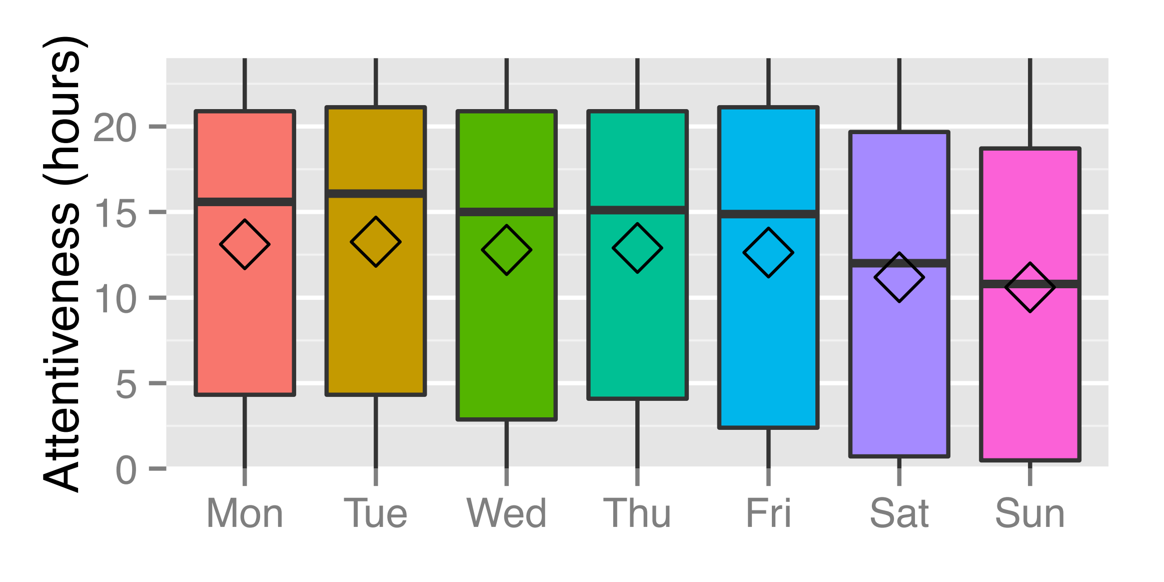
Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**.

- Stepwise iteration through all sensors
- Classification
  - *Attentive* vs *non-attentive*: triage messages within 2.08min
- Random Forrest: 79.29% accuracy and $\kappa = .586$.
  - Precision of .771
  - Recall of .828
- 86,400 states per day

Predicting Attentiveness for Every Minute of the Day

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**.

Attentiveness Throughout the Week



F (6, 20802) = 41.07, p<001

Weekdays vs. Weekends: p<.001

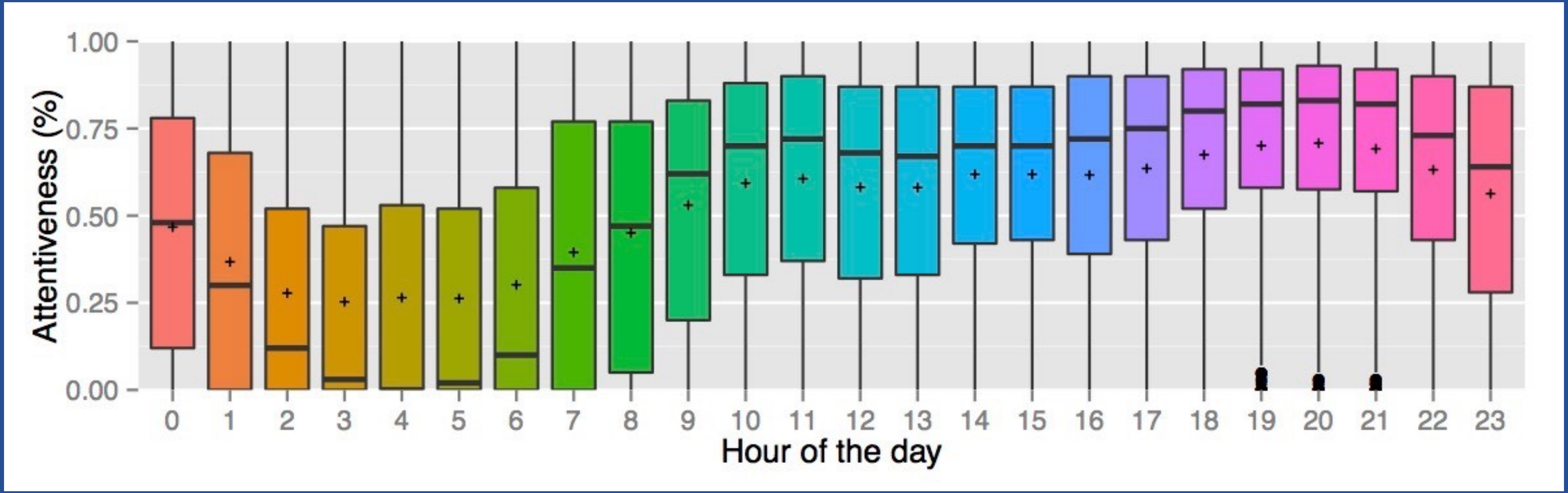
Week: 62%-67% attentive

Weekends: 45%-50%

Average attentiveness by day

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI '15**.

Attentiveness Throughout the Day



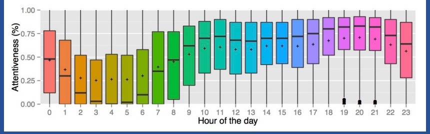
- Average attentiveness by hour

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**.

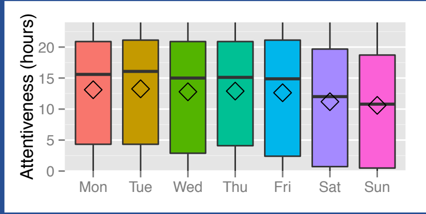
High Attentiveness to Mobile Messaging

- *Attentive* for on avg. 50.5% (SD=14.6%) of the full day
- *i.e.*

12.1 HOURS PER DAY



84.8 HOURS PER WEEK



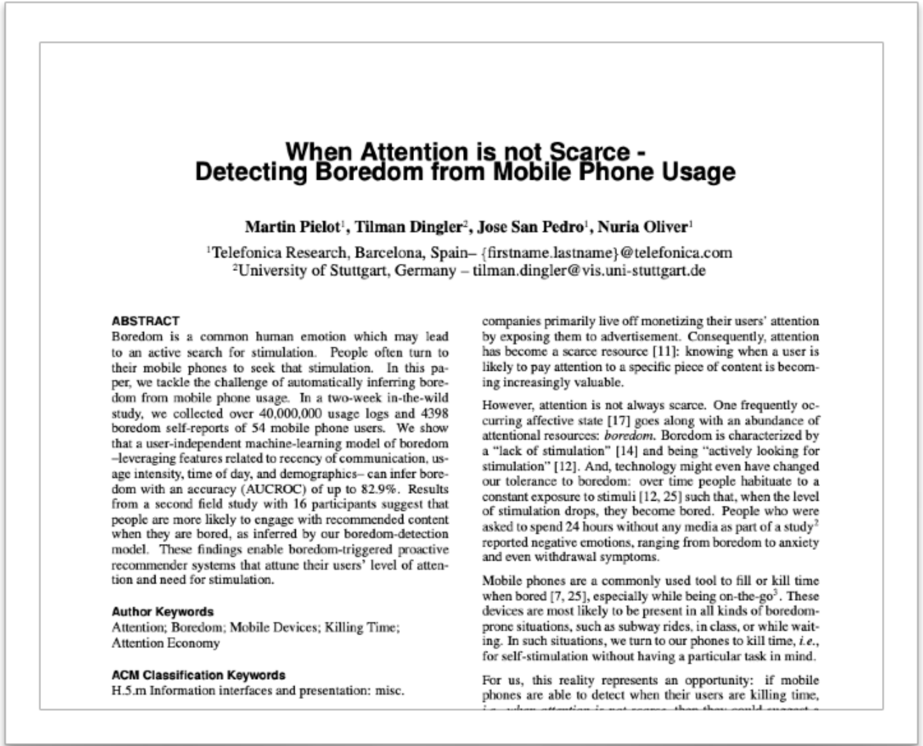
75.8% OF THE WAKING HOURS

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**. http://dx.doi.org/10.1145/2785830.2785840

High Attentiveness to Mobile Messaging

- People are **attentive** to messages **12.1 hours of the day**
- Attentiveness is **higher during the week** than on the weekend
- People are **more attentive during the evening**
- When being inattentive, people **return** to attentive states **within 1-5 minutes** in the majority (75% quantile) of the cases

Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In Proc. of **MobileHCI ’15**. http://dx.doi.org/10.1145/2785830.2785840



When Attention is not Scarce -
Detecting Boredom from Mobile Phone Usage

Martin Pielot[1], Tilman Dingler[2], Jose San Pedro[1], Nuria Oliver[1]

[1]Telefonica Research, Barcelona, Spain– {firstname.lastname}@telefonica.com
[2]University of Stuttgart, Germany – tilman.dingler@vis.uni-stuttgart.de

ABSTRACT
Boredom is a common human emotion which may lead to an active search for stimulation. People often turn to their mobile phones to seek that stimulation. In this paper, we tackle the challenge of automatically inferring boredom from mobile phone usage. In a two-week in-the-wild study, we collected over 40,000,000 usage logs and 4398 boredom self-reports of 54 mobile phone users. We show that a user-independent machine-learning model of boredom –leveraging features related to recency of communication, usage intensity, time of day, and demographics– can infer boredom with an accuracy (AUCROC) of up to 82.9%. Results from a second field study with 16 participants suggest that people are more likely to engage with recommended content when they are bored, as inferred by our boredom-detection model. These findings enable boredom-triggered proactive recommender systems that attune their users' level of attention and need for stimulation.

Author Keywords
Attention; Boredom; Mobile Devices; Killing Time; Attention Economy

ACM Classification Keywords
H.5.m Information interfaces and presentation: misc.

companies primarily live off monetizing their users' attention by exposing them to advertisement. Consequently, attention has become a scarce resource [11]: knowing when a user is likely to pay attention to a specific piece of content is becoming increasingly valuable.

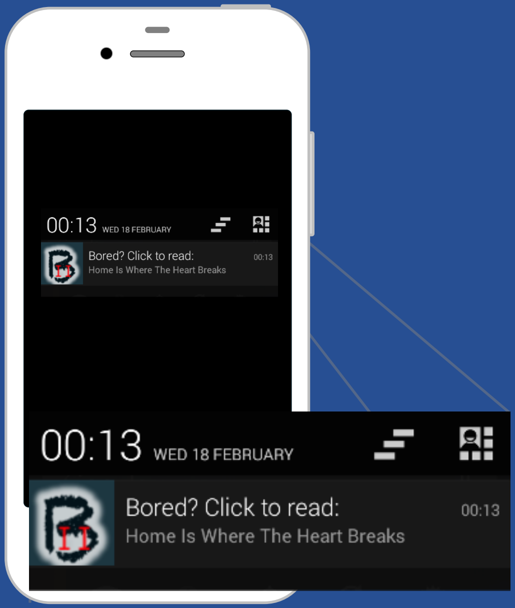
However, attention is not always scarce. One frequently occurring affective state [17] goes along with an abundance of attentional resources: *boredom*. Boredom is characterized by a "lack of stimulation" [14] and being "actively looking for stimulation" [12]. And, technology might even have changed our tolerance to boredom: over time people habituate to a constant exposure to stimuli [12, 25] such that, when the level of stimulation drops, they become bored. People who were asked to spend 24 hours without any media as part of a study[2] reported negative emotions, ranging from boredom to anxiety and even withdrawal symptoms.

Mobile phones are a commonly used tool to fill or kill time when bored [7, 25], especially while being on-the-go[3]. These devices are most likely to be present in all kinds of boredom-prone situations, such as subway rides, in class, or while waiting. In such situations, we turn to our phones to kill time, *i.e.*, for self-stimulation without having a particular task in mind.

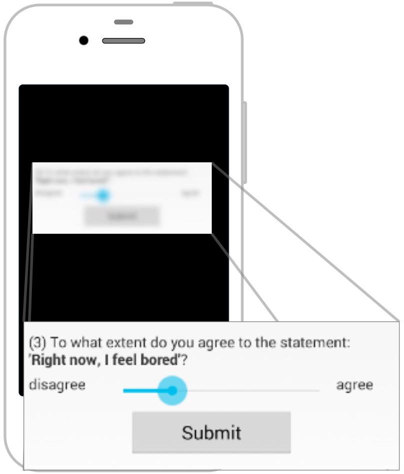
For us, this reality represents an opportunity: if mobile phones are able to detect when their users are killing time,

Proof-of-Concept Study

Using Boredom to Predict Opportune Moments for Content Delivery



M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver. When attention is not scarce - detecting boredom from mobile phone usage. **UbiComp '15**.



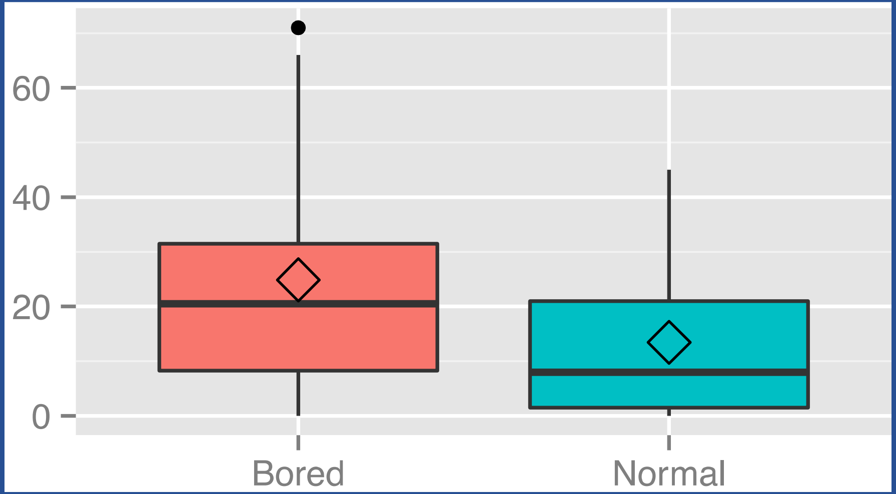
Category	Example Feature	Explanation
Context	Semantic Location	Home, work, other, unknown
Demographics	Age, gender	38, female
Last Communication Activity	Time last incoming call	Time passed since somebody called the participants
Usage (intensity)	Bytes received	Number of bytes downloaded in the last 5 minutes
Usage (externally triggered)	Number of notifications	Number of notifications received in the last 5 minutes
Usage (idling)	Number of apps	Number of apps launched in the last 5 minutes
Usage (type)	Most used app	App used for the most time in the last 5 minutes.

Boredom can be detected from phone-usage patterns with an accuracy of approx. 75% to 83% AUCROC

We found subjective boredom to be related to regency of communication, phase of the day demographics, intensity and type of phone usage, and type of used apps

M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver. When attention is not scarce - detecting boredom from mobile phone usage. **UbiComp '15**. http://dx.doi.org/10.1145/2750858.2804252

People Are More Likely to **Click**



% of clicked notifications (Mdn)

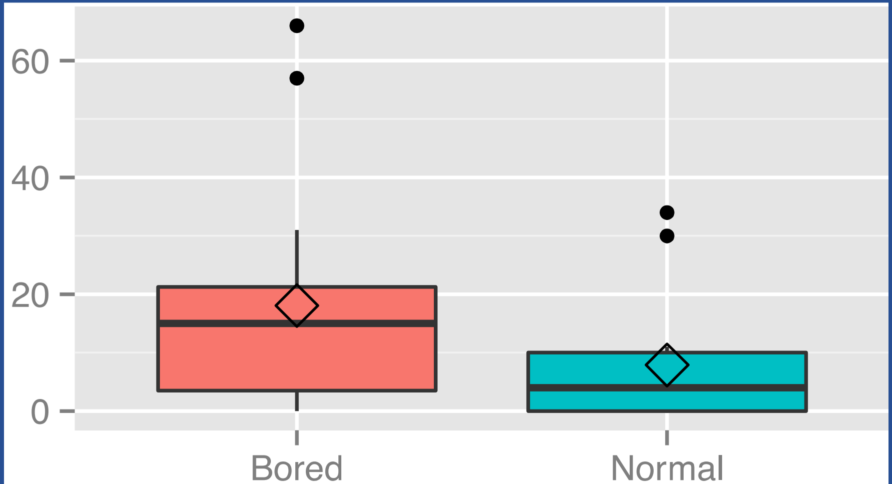
8% when **not bored**
20.5% when **bored**
(as inferred by the model)

Difference significant
z = -2.102, p = .018

Large effect
r = -.543

M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver. When attention is not scarce - detecting boredom from mobile phone usage. **UbiComp '15**. http://dx.doi.org/10.1145/2750858.2804252

People Are More Likely to Stick



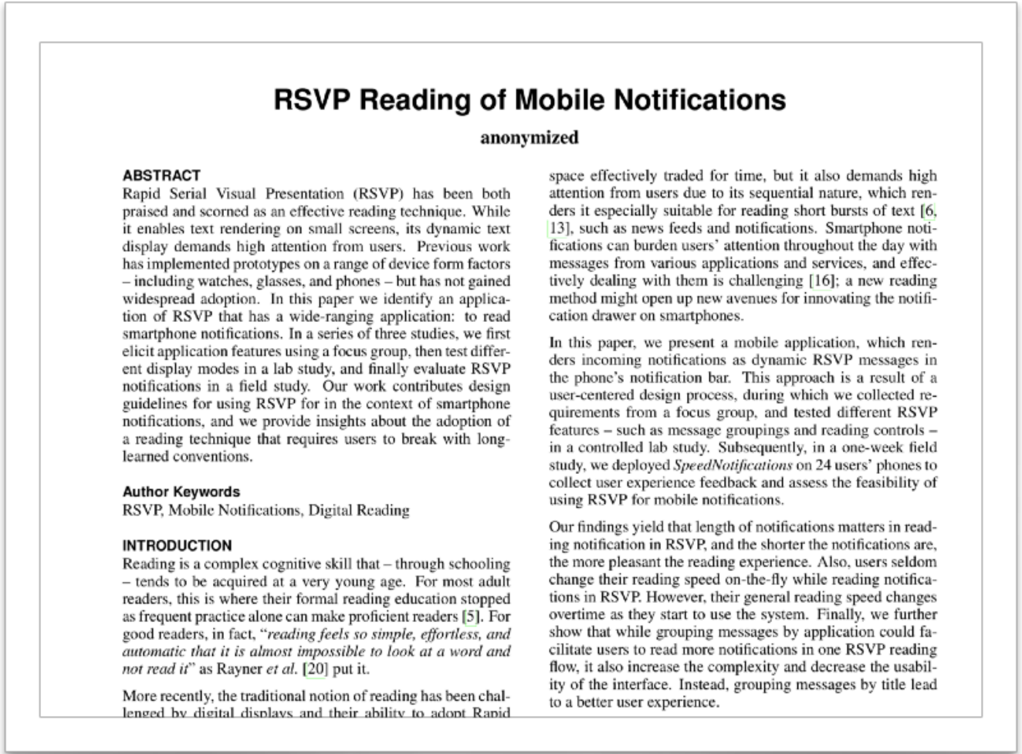
% of clicks where people spent more than 30 sec reading (Mdn)

4% when **not bored**
15% when **bored**
(as inferred by the model)

Difference significant
z = -2.102, p = .018

Large effect
r = -.511

M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver. When attention is not scarce - detecting boredom from mobile phone usage.
UbiComp '15. http://dx.doi.org/10.1145/2750858.2804252



RSVP Reading of Mobile Notifications

anonymized

ABSTRACT

Rapid Serial Visual Presentation (RSVP) has been both praised and scorned as an effective reading technique. While it enables text rendering on small screens, its dynamic text display demands high attention from users. Previous work has implemented prototypes on a range of device form factors – including watches, glasses, and phones – but has not gained widespread adoption. In this paper we identify an application of RSVP that has a wide-ranging application: to read smartphone notifications. In a series of three studies, we first elicit application features using a focus group, then test different display modes in a lab study, and finally evaluate RSVP notifications in a field study. Our work contributes design guidelines for using RSVP for in the context of smartphone notifications, and we provide insights about the adoption of a reading technique that requires users to break with long-learned conventions.

Author Keywords

RSVP, Mobile Notifications, Digital Reading

INTRODUCTION

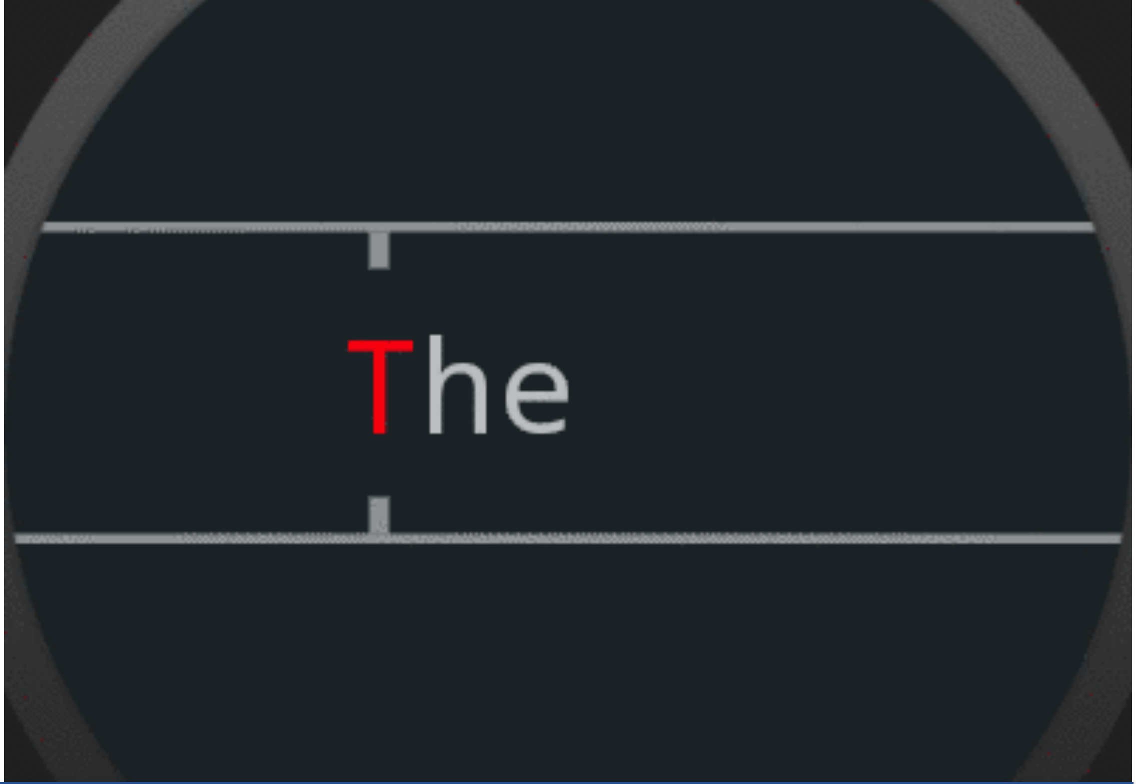
Reading is a complex cognitive skill that – through schooling – tends to be acquired at a very young age. For most adult readers, this is where their formal reading education stopped as frequent practice alone can make proficient readers [5]. For good readers, in fact, "*reading feels so simple, effortless, and automatic that it is almost impossible to look at a word and not read it*" as Rayner *et al.* [20] put it.

More recently, the traditional notion of reading has been challenged by digital displays and their ability to adopt Rapid space effectively traded for time, but it also demands high attention from users due to its sequential nature, which renders it especially suitable for reading short bursts of text [6, 13], such as news feeds and notifications. Smartphone notifications can burden users' attention throughout the day with messages from various applications and services, and effectively dealing with them is challenging [16]; a new reading method might open up new avenues for innovating the notification drawer on smartphones.

In this paper, we present a mobile application, which renders incoming notifications as dynamic RSVP messages in the phone's notification bar. This approach is a result of a user-centered design process, during which we collected requirements from a focus group, and tested different RSVP features – such as message groupings and reading controls – in a controlled lab study. Subsequently, in a one-week field study, we deployed *SpeedNotifications* on 24 users' phones to collect user experience feedback and assess the feasibility of using RSVP for mobile notifications.

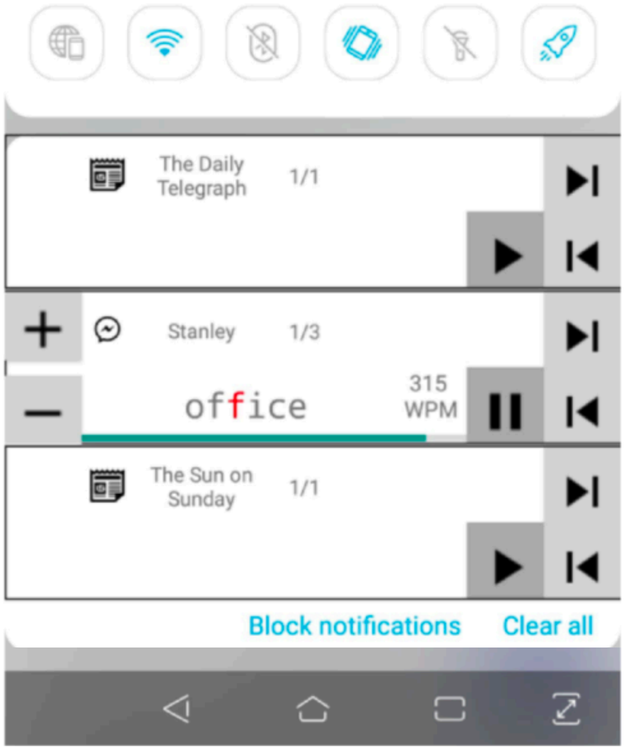
Our findings yield that length of notifications matters in reading notification in RSVP, and the shorter the notifications are, the more pleasant the reading experience. Also, users seldom change their reading speed on-the-fly while reading notifications in RSVP. However, their general reading speed changes overtime as they start to use the system. Finally, we further show that while grouping messages by application could facilitate users to read more notifications in one RSVP reading flow, it also increase the complexity and decrease the usability of the interface. Instead, grouping messages by title lead to a better user experience.

Experience Using a Prototype



Rapid Serial Visual Presentation

- 1-week field study
- 24 participants (12 female)
- 2 between-group conditions
  - Group notification by app
  - Group notification by title
- Measures
  - App logs
  - Interview



Re-designing Reading of Notifications

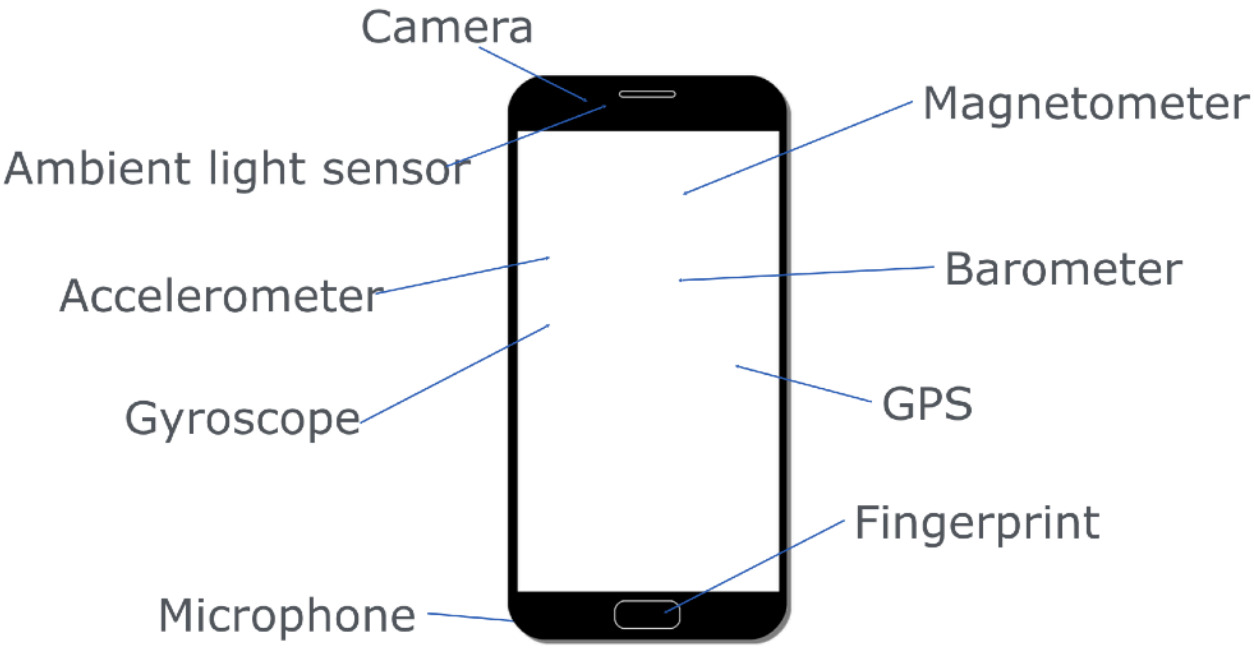
Hao-Ping Lee, Benjamin Tag, Qiushi Zhou, Vassilis Kostakos, Tilman Dingler. "RSVP Reading of Mobile Notifications." **Under submission**, 2019.

- Ethnography / Observations
- Diary Studies
- Experience Sampling
- Experiments

Field Study Methods

Student Experience Survey

https://ses.unimelb.edu.au

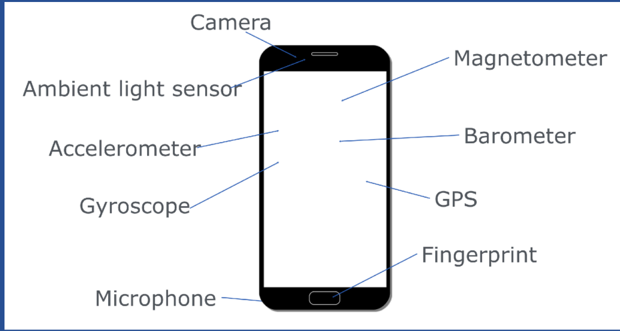


Sensor-Based Observation

Sensor-based Observation

- Sensors
  - Provide a truly objective measure
  - Sensors are everywhere (think about mobile phones, sensors in home automation, etc.)
  - Sensors emerge into the background (i.e., they are pervasive), thus they are quite unobtrusive

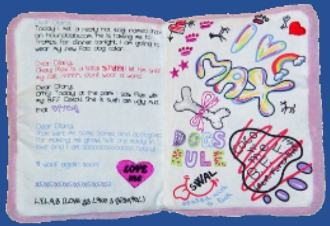
- Examples
  - GPS (location, orientation, …)
  - Accelerometer (activity, device posture, …)
  - Microphone (environmental noise level, …)
  - Camera (is the user looking on the display, …)
  - Key-/Touch-Logger (interaction patterns, …)
  - Device State (which App is running, charging, …)
  - Etc.



Diary Studies

Diary Studies

- Asks people to keep a diary, or journal, of their interactions with a computer system, any significant events or problems during their use of a system, or other aspects of their working life.
- A diary typically asks a user to record the date and time of an event, where they are, information about the event of significance, and ratings about how they feel, etc.
- An interesting alternative for making diary entries is to give users a tape recorder (or a mobile phone...) and a list of questions, so that users don't need to write things down as they encounter them.



Conducting Diary Studies

- Provide participants with pen and paper to record their entries
- Provide participants with alternative means to record data
  - eMail address
  - voice recorder
  - photo camera
  - camcorder
- Provide a semi-structured format for recording entries
  - Highly structured formats can prevent recording unanticipated events
  - Unstructured formats might confuse: „How should I record my events?"
- Diary study should be followed up by in-depth interviews with participants

Pros and cons of diary studies

- Pros
  - Cheaper than location-based ethnography
  - Long term observation
  - Very good to investigate context of use
- Cons
  - Depends on the users motivation (might decrease early)
  - Hardly to find out, if diary is complete or what is missing (not very reliable)

Context-Informed Scheduling and Analysis: Improving Accuracy of Mobile Self-Reports

Niels van Berkel
Jorge Goncalves
Peter Koval
first.last@unimelb.edu.au
The University of Melbourne, Australia

Simo Hosio
first.last@oulu.fi
University of Oulu, Finland

Tilman Dingler
first.last@unimelb.edu.au
The University of Melbourne, Australia

Denzil Ferreira
first.last@oulu.fi
University of Oulu, Finland

Vassilis Kostakos
first.last@unimelb.edu.au
The University of Melbourne, Australia

ABSTRACT

Mobile self-reports are a popular technique to collect participant labelled data in the wild. While literature has focused on increasing participant compliance to self-report questionnaires, relatively little work has assessed response accuracy. In this paper, we investigate how participant context can affect response accuracy and help identify strategies to improve the accuracy of mobile self-report data. In a 3-week study we collect over 2,500 questionnaires containing both verifiable and non-verifiable questions. We find that response accuracy is higher for questionnaires that arrive when the phone is not in ongoing or very recent use. Furthermore, our results show that long completion times are an indicator of a lower accuracy. Using contextual mechanisms readily available on smartphones, we are able to explain up to 13% of the variance in participant accuracy. We offer actionable recommendations to assist researchers in their future deployments of mobile self-report studies.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; *Ubiquitous and mobile computing design and evaluation methods*;

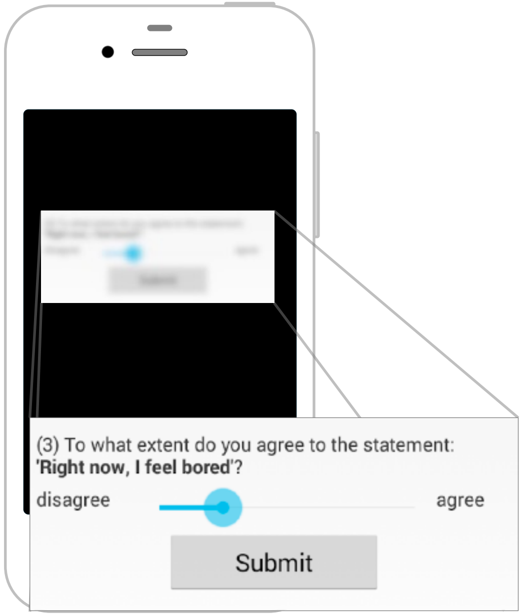
ACM Reference Format:
Niels van Berkel, Jorge Goncalves, Peter Koval, Simo Hosio, Tilman Dingler, Denzil Ferreira, and Vassilis Kostakos. 2019. Context-Informed Scheduling and Analysis: Improving Accuracy of Mobile Self-Reports. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300281

1 INTRODUCTION

The Experience Sampling Method (ESM) [33], also known as Ecological Momentary Assessment [7, 50], is widely used to obtain *in situ* data on a variety of research topics, including affective state [47], activities [15], and technology usage [12, 43]. Participants in ESM studies answer a set of questions, *i.e.* self-reports, throughout the day, typically for 1-3 weeks [16]. As researchers rely on the input of study participants rather than direct observation, attaining a sufficient level of accuracy and frequency of completed self-reports is crucial. Previous literature has focused on answer frequency, for example by timing questionnaires based on participant context [45] or by increasing motivation [8, 24]. However, assessing the accuracy of ESM responses has received little attention in the literature, despite the considerable implications for study results. Researchers tend to assume that

The Experience Sampling Method (ESM)

- Subjective ground truth collection
- Mini-surveys triggered throughout the day
  - Time-dependent trigger
  - Event-triggered
- Pro
  - In-situ
  - Reduced recall bias
- Con
  - Frequency and extent need to be carefully considered



The Experience Sampling Method (ESM)

- Experiments are (probably the only reliable) means to find the answer
  - They can **isolate cause and effect**!
  - Knowing cause and effect allows informed design

Experiments

Experiments in Science

Hypothesis: Meerkats fall asleep when given whisky.

A

B

Experimental Design I

- Independent Variables
  - The manipulated aspects: whisky / no whisky
- Dependent Variables
  - The observed aspects: meerkat sleepiness
- Conditions
  - The levels of the independent variables {whisky, no-whisky, (different dosages…)}
  - Control Condition (absence of suspected cause: no whisky)
  - Experimental Conditions (presence of suspected cause: whisky)

Experimental Design II

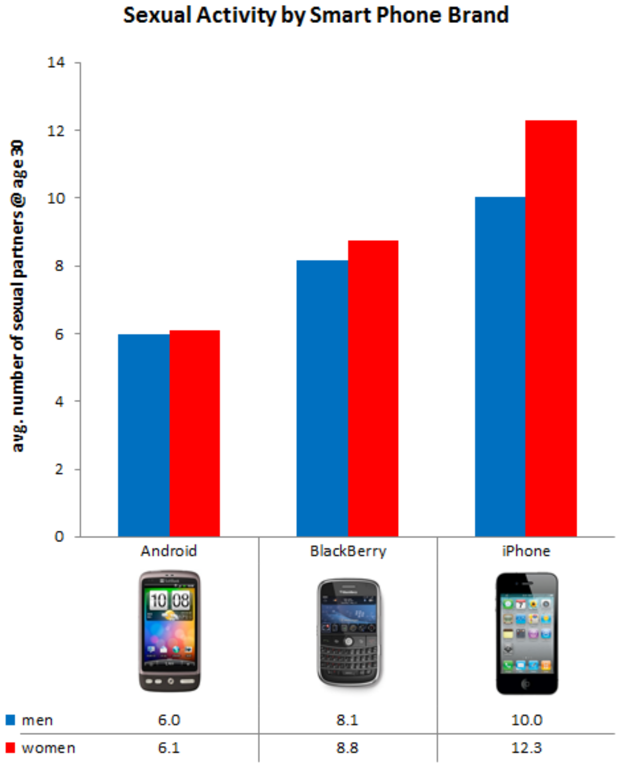
- Repeated measures/**within subjects**
  - All participants (i.e., meerkat) are assigned to all conditions
- or independent measures/**between groups**
  - Participants (i.e., meerkat) are assigned to one condition only
- Experimental design vs. quasi-experimental design

Dependent Variables
“what we measure…”

- The results are commonly referred to as scores
- They can be measured in different scales:
  - Categorial or Nominal (drunk, sober, sleepy)
  - Ordinal (low, medium, high)
  - Interval (blood alcohol concentration)
  - Ratio (time in ms)
- The difficulty is to find a valid dependent variable

The independent Variable
“what we change…”

- How to manipulate a single aspect only?
- In theory: by keeping all other factors stable
  - Environment, weather, intelligence, mood, …
  - BUT, people, situations, … are never identical!



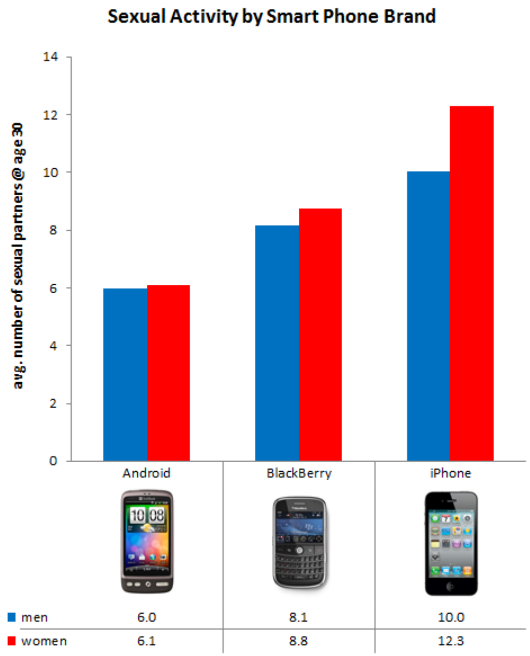
How would you **design an experiment** from this?

Will buying a new phone change my sexual activity?

OkCupid, 2010

10min Group Exercise
Experimental Design

- Independent Variables
  - The manipulated aspects: whisky / no whisky
- Dependent Variables
  - The observed aspects: meerkat sleepiness
- Conditions
  - The levels of the independent variables {whisky, no-whisky, (different dosages…)}
  - Control Condition (absence of suspected cause: no whisky)
  - Experimental Conditions (presence of suspected cause: whisky)
- In-between vs. between-groups Design

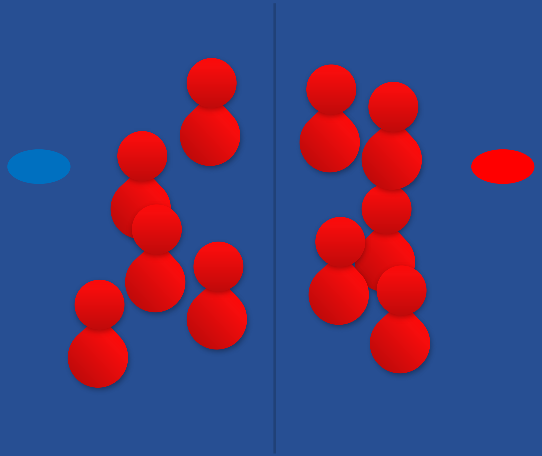


Quasi- and True Experimental Designs

- True experiment: full control over independent variable
  - But, independent variables cannot always be controlled:
    - Gender, Age
    - Behavior / actions that are not recommended
- This is then called quasi-experimental design
- But
  - Are age-typical impairments the cause?
  - Or e.g. the lack of life-long experience with computers?
    (so age is not really the cause)
- Limits the conclusions we can draw from the study

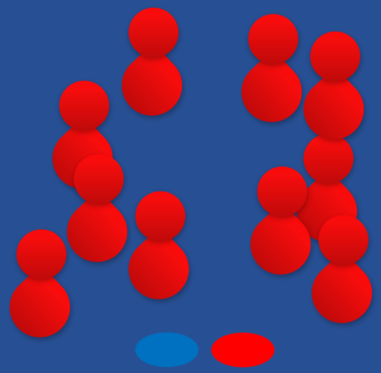
Experimental designs II

- Between-groups – independent measures
  - Also called independent-measures design
  - Participants are assigned to one condition only



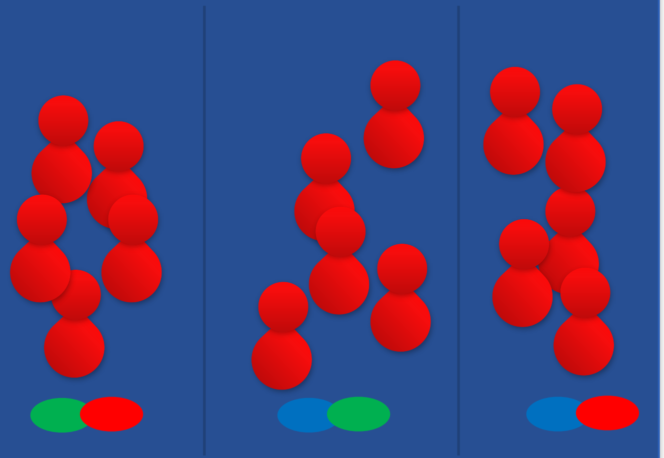
Experimental designs II

- Within-subjects - repeated-measures
  - Also called repeated-measures design
  - Participants are assigned to all conditions
  - Order must be counter-balanced or randomized



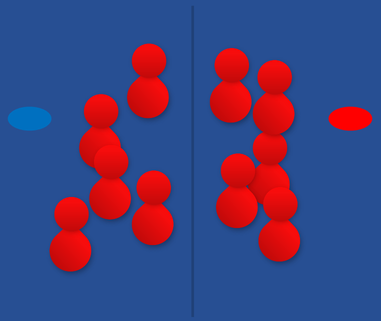
Experimental designs II

- Hybrid (mixed) designs also possible



Independent-measures design (between groups)

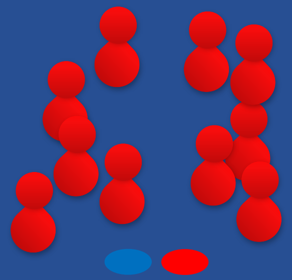
- Advantages
  - Simplicity
  - Less chance of practice or fatigue effects
  - Useful when it is impossible for an individual to participate in all conditions
- Disadvantages
  - Expense (time, effort, and number of participants)
  - Insensitiveness to experimental manipulations



Repeated-measures design (within groups)

- Advantages
  - Economy
  - Sensitiveness
  - Cancelling out individual differences

- Disadvantages
  - Carry-over effects from previous conditions
  - Conditions need to be reversible



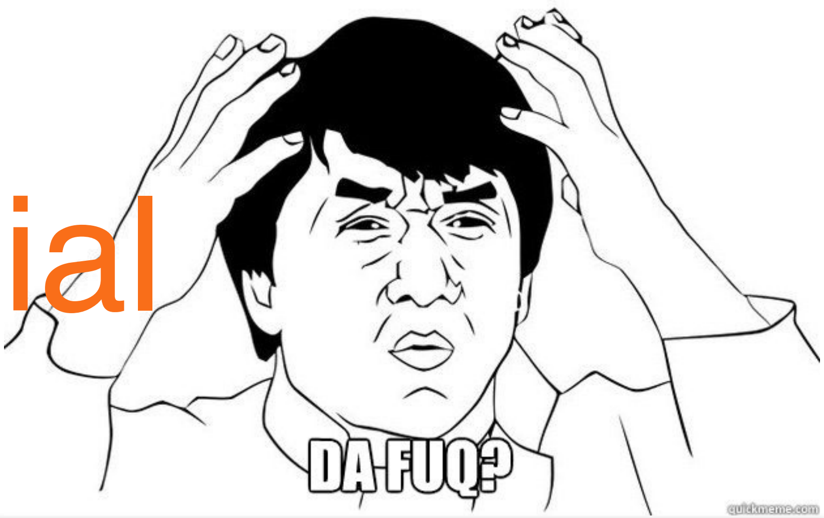
So which design should I use?

- Repeated-measures (within groups) over independent-measures (between groups)
  - Differences between participants are cancelled out
  - Less participants are required
- Experimental design over quasi-experimental design
  - Since we can draw stronger conclusions
  - Be aware of side effects when using quasi-experimental designs

Descriptive

vs

Inferential



Statistics

My brother’s girlfriend from England says she can taste whether milk or tea was poured in first…

Image by Xavier Snelgrove from http://en.wikipedia.org/wiki/File:Milk_clouds_in_tea.jpeg

Statistical Significance

- Example: tea with milk
  - My brother's girlfriend from England says she can taste whether milk or tea was poured in first…
- Probability for consecutive correct guessing
  - 1x, 2x, 3x, 4x, 5x, …
  - 0.5, 0.25, 0.125, 0.0625, 0.03125, …
- Typical levels of significance
  - $p < .05$ (Fisher-Criterion)
  - $p < .01$
  - $p < .001$
  - => 5 correct guesses in a row < Fisher-Criterion

Answering Research Questions

Inferential

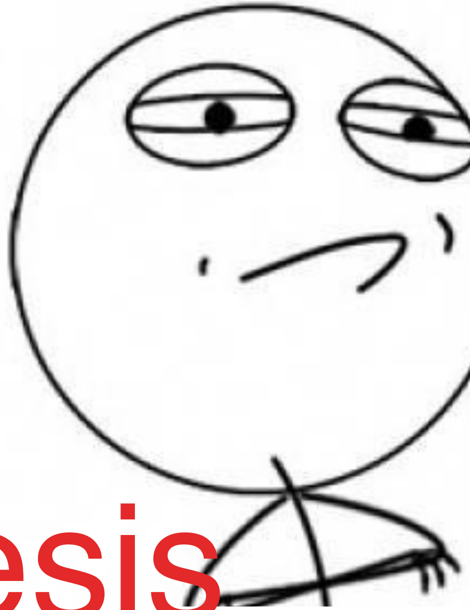
- Testing Experimental Hypotheses
- Working with probabilities
  (chance results?)
- Fisher Criterion: 95%
- Sources of variation

$$TestStatistic = \frac{Systematic}{Unsystematic}$$

- if p<=0.5

Rejecting the

null hypothesis



null hypothesis

noun

(in a statistical test) the hypothesis that there is no significant difference between specified populations, any observed difference being due to sampling or experimental error.

Photo by amk713 http://www.flickr.com/photos/nylffn/3505943314 (CC BY-NC-ND 2.0)

Informed Consent

“Primum non nocere”/”First, do no harm” (Thomas Sydenham)

- An Informed Consent is **necessary**
  - when human subjects participate in studies
    - Take part in a study
    - Are interviewed
    - …
  - when personally identifiable information is collected
    - … which is information that can
    - uniquely identify, contact, or locate a single person, or
    - uniquely identify a single individual with other sources

Informed Consent

- Background
  - context in which the research takes place?
  - what do participants have to expect?
- Data Collection
  - what kind of data will be logged?
  - how has access and how will it be secured?
  - what is going to be reported?
- Legal Rights
  - Information participants that
    - They do not have to e.g. answer questions if they don’t want to
    - They can cancel the experiment at any time without explanation

Ethical Considerations

"One should treat others as one would like others to treat oneself"
The "Golden Rule"

- Deception
  - Consider whether it is ethical to not inform participants about the real context of the experiment
- Debriefing
  - After the experiment, answer the participants questions
- Confidentially
  - Keep information confidential at any time!
- Protection from physical and psychological harm
  - Do not physically or psychologically harm your participants

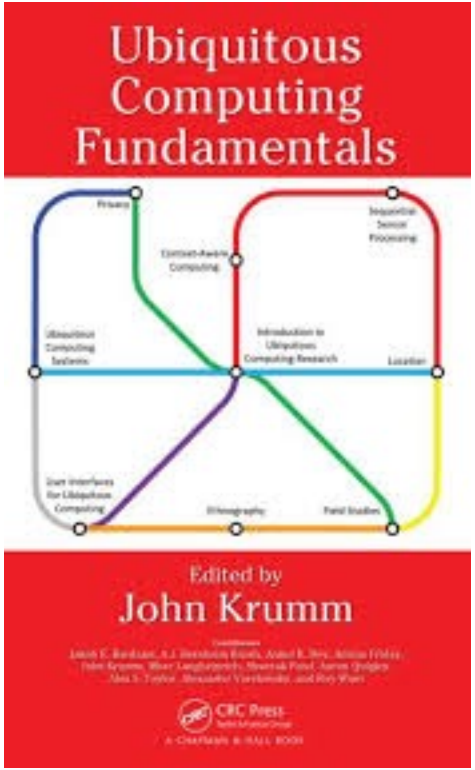
- **Field studies** are used to
    - Study user behaviour
    - Conduct proof-of-concept studies
    - Experience using a prototype

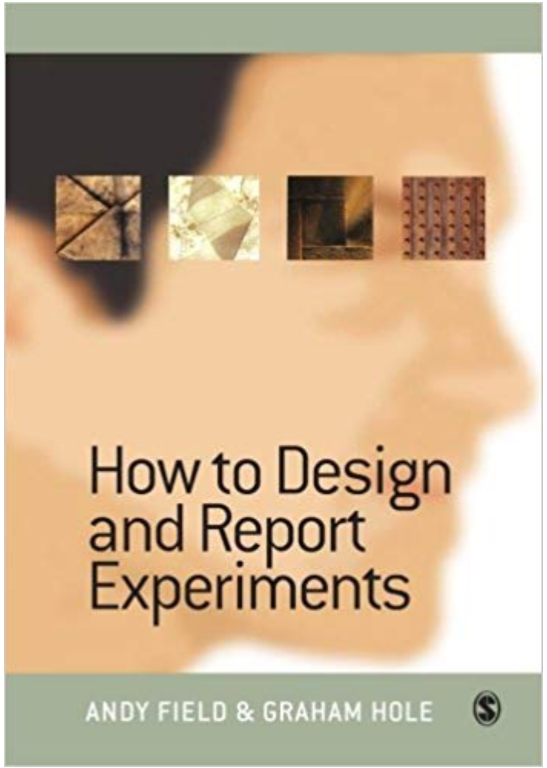
- **Methods** include
    - Ethnography/Observations
    - Diary studies
    - Experience Sampling Method
    - Experiments

- **Statistics** help us to make **inferences**
    - Use descriptive statistics to summarize observed outcomes
    - Use inferential statistics to calculate probability that effect was no coincidence

- Results are **never true** in a sense of **being 100% correct**!
    - The world is messy, get used to it, have fun with it!

Conclusion





References and Further Readings