INFOMSCIP 2019-2020
lectures 10&11
Oct 7&10, 2019

Many of these slides are based on the following two sources:

- J. Borchers' course on "Current Topics in Media Computing and Human-Computer Interaction" at RWTH Aachen
  https://hci.rwth-aachen.de/cthci
- S. MacKenzie's course on "Empirical research methods in human-computer interaction" at ACM CHI 2018
  http://www.yorku.ca/mack/CourseNotes.pdf
  (which in turn is based on his book "Human Computer Interaction: An Empirical Research Perspective")

Both are highly recommended if you want to dig deeper into the topic (the course by J. Borcher's also has recordings of his lectures).

I also took some inspiration from the slides of the INFOARM course and other sources, some of which are referenced on the slides.

Scientific perspectives on GMT (INFOMSCIP)

Recap & intro for user-related research

- Recap & the bigger picture
  Putting experimental and user-related research in context
- Research questions
  What is it and why do we need it

Branches of science / scientific fields / scientific disciplines:

Formal sciences: the study of mathematics and logic.

Natural sciences: the study of natural phenomena.

Social sciences: the study of human behavior and societies.

Natural, social, and formal science make up the **fundamental sciences**, which form the basis of **interdisciplinary** and **applied sciences** such as engineering and medicine.

From Wikipedia.org ("Branches of science")

Computer science: part of applied sciences
Theoretical computer science: part of formal sciences

⇒ Different branches use different scientific methods

⇒ Computer science applies methods from all of them (and sometimes combines them)

From lecture 1 (intro)

Common research approaches in GMT

Fundamental research:

Not related to specific data, nor to users.

Provides answers to universal questions within well-known and well-accepted scientific frameworks.

Experimental research:

Done on a data set that can come from the real world (by measurements) or that may be generated (synthetic data).

Answers to questions cannot be universal: whatever is observed is observed for the tested data only, and not for all conceivable data.

The research question itself can be a theoretical question or an applied question.

User study research:

Studies or observes humans and their behavior.

Can also be rather fundamental or more applied.

In all cases, answers to user study research questions tell us something about users, and not about (non-user) data or abstract frameworks.

From lecture 1 (intro)

Common research approaches in GMT

Fundamental research
Not related to specific data, nor to users.
Answers to universal questions within
well-known/accepted scientific frameworks.

The "framework is the **data**"
(and synthetic and measured data
can be helpful, too, e.g., to verify
a framework)

Experimental research
Done on a data set from
the real world (by measurements)
or generated (synthetic data).
Answers questions about the test data.
Can be a theoretical/fundamental or applied.

Needs **data**
(created or existing)

User study research
Studies or observes humans and their behavior.
Answers questions about users.
Can also be fundamental or applied.

Needs **data**
(from users)

*Difficult, because humans
are different. And difficult.*

Common research approaches in GMT

Fundamental research
Not related to specific data, nor to users.
Answers to universal questions within well-known/accepted scientific frameworks.

Even for fundamental research, **humans** can be relevant

- Judging outcome (or defining it; researchers are humans, too), e.g., quality, difficulty (puzzle games)
- Measure experience, enjoyment, … (e.g., in relation to difficulty)

Experimental research
Done on a data set from
the real world (by measurements)
or generated (synthetic data).
Answers questions about the test data.
Can be a theoretical/fundamental or applied.

Measured **data** can be **about users and their behavior** (or influenced or created by them).

User study research
Studies or observes **humans** and their behavior.
Answers questions about users.
Can also be fundamental or applied.

Users are needed to observe and gather data.

Common research approaches in GMT

Fundamental research
Not related to specific data, nor to users.
Answers to universal questions within
well-known/accepted scientific frameworks.

Needs **measures**, e.g.,
to define frameworks,
to compare frameworks,
to verify/proof frameworks,
…

Experimental research
Done on a data set from
the real world (by measurements)
or generated (synthetic data).
Answers questions about the test data.
Can be a theoretical/fundamental or applied.

Needs **measures**, e.g.,
to quantify real-world data,
to create and test synthetic data,
…

User study research
Studies or observes humans and their behavior.
Answers questions about users.
Can also be fundamental or applied.

Needs **measures**, e.g.,
to quantify human behavior,
to estimate humans' opinions,
…

Common research approaches in GMT

Fundamental research
Not related to specific data, nor to users.
Answers to universal questions within well-known/accepted scientific frameworks.

Frameworks needs to be well-known and well-accepted (guaranteed via peer-review).
Definition of framework and measures are important!

Experimental research
Done on a data set from
the real world (by measurements)
or generated (synthetic data).
Answers questions about the test data.
Can be a theoretical/fundamental or applied.

User study research
Studies or observes humans and their behavior.
Answers questions about users.
Can also be fundamental or applied.

The goal is also to gain "universal" knowledge.
Selection of good data and measures are important!

Common research approaches in GMT

Fundamental research
Not related to specific data, nor to users.
Answers to universal questions within well-known/accepted scientific frameworks.

Experimental research
Done on a data set from
the real world (by measurements)
or generated (synthetic data).
Answers questions about the test data.
Can be a theoretical/fundamental or applied.

User study research
Studies or observes humans and their behavior.
Answers questions about users.
Can also be fundamental or applied.

Bottom line:

We **always** must make decisions about framework/context, data, measures, …
that **impact our results**.

Differences exist between fields, but there are also **commonalities and similarities.**

Science is about creating 'new knowledge'.
Engineering & design are about creating 'new things'.

In game & media technology, we often want to create new technology.
Human-related research in GMT helps us achieve this in two ways:

When developing a new technology (or concept, design, …),
research can answer questions such as:

- **Which abilities and limitations do humans possess that should be taken into account in the new design?**

Examples:

- What vibration patterns are recognizable by average humans? (And under what circumstances?)
- How many different vibration patterns can people remember?

When verifying if a new technology (or concept, design, …) works,
research can answer questions such as:

- **Does it meet its design goals?**

Examples:

- Does a new input device result in time savings/likeability?

When you do a research project, people often ask you:

What is your research questions?

To answer this, you first need to know:

What is a research question?

A **research question** is the objective of a study or a problem to be solved through research. Choosing a research question is an essential element of both quantitative and qualitative research.

The research question can take different forms depending on the type of research.

From Wikipedia ("Research question")

Note: There is no single agreement on this across all sciences (or even between people within one branch of science).

Not all research needs a research question.

But they are important if not essential in empirical research.

Quantitative and objective are related aspects, but not the same.
The same goes for qualitative and subjective.

	Quantitative	Qualitative
Objective	"The chip of my computer is 2 GHz." "It took 30 sec to solve the task with this approach."	"Yes, I own a computer." "Yes, I solved the task with this approach."
Subjective	"On a scale from 1-10, my computer scores 7 in terms of its ease of use." "In terms of speed, I would rate this approach as 7 on a scale from 1-10."	"I think computers are too expensive." "The approach allowed me to solve the task quite fast."

From https://www.userfocus.co.uk/articles/datathink.html (blue parts have been added)

In-class discussion: Can we come up with a good specification of essential or helpful characteristics for a good research question?

Top 9 Main Characteristics of Science:

1. Objectivity
2. Verifiability
3. Ethical Neutrality
4. Systematic Exploration
5. Reliability
6. Precision
7. Accuracy
8. Abstractness
9. Predictability

Source: http://www.yourarticlelibrary.com/science/top-9-main-characteristics-of-science-explained/35060

Good research questions should …

… follow ethical standards.

… be testable.

… allow for reproducibility, repeatability.

… generate new, relevant knowledge.

…

Empirical user studies

- The empirical approach
  What is empirical research?
- Research questions & hypothesis
  What & why. Variables & validity of an experiment.
- Study design
  Subjects, environment & other contexts,
  within- versus between-subjects designs,
  order effects.

Empirical research is research using empirical evidence.
It is a way of gaining knowledge by means of direct and indirect observation or experience.

From Wikipedia ("Empirical research")

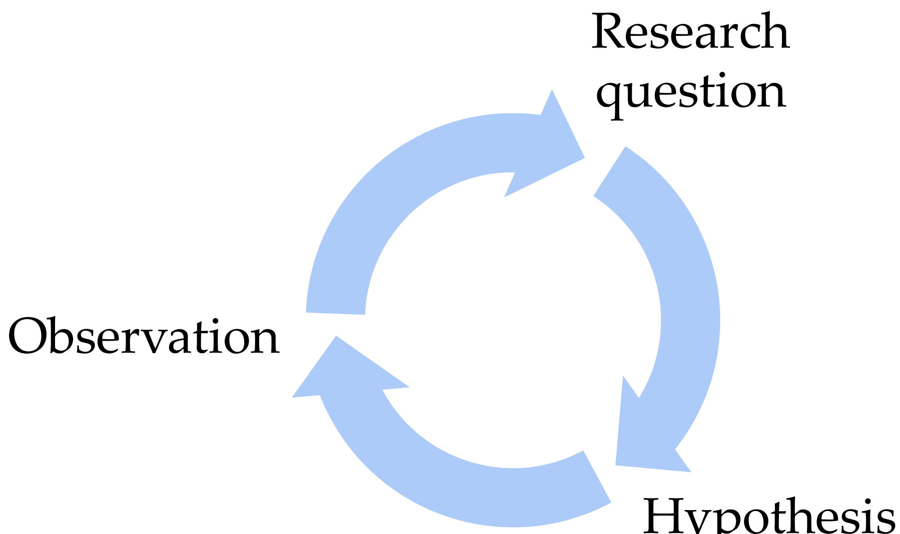
Empirical evidence is the information received by means of the senses, particularly by observation and documentation of patterns and behavior through experimentation.

From Wikipedia ("Empirical evidence")

Quote from a preceding slide about *experimental research*:
Answers to questions cannot be universal: whatever is observed is observed for the tested data only, and not for all conceivable data.

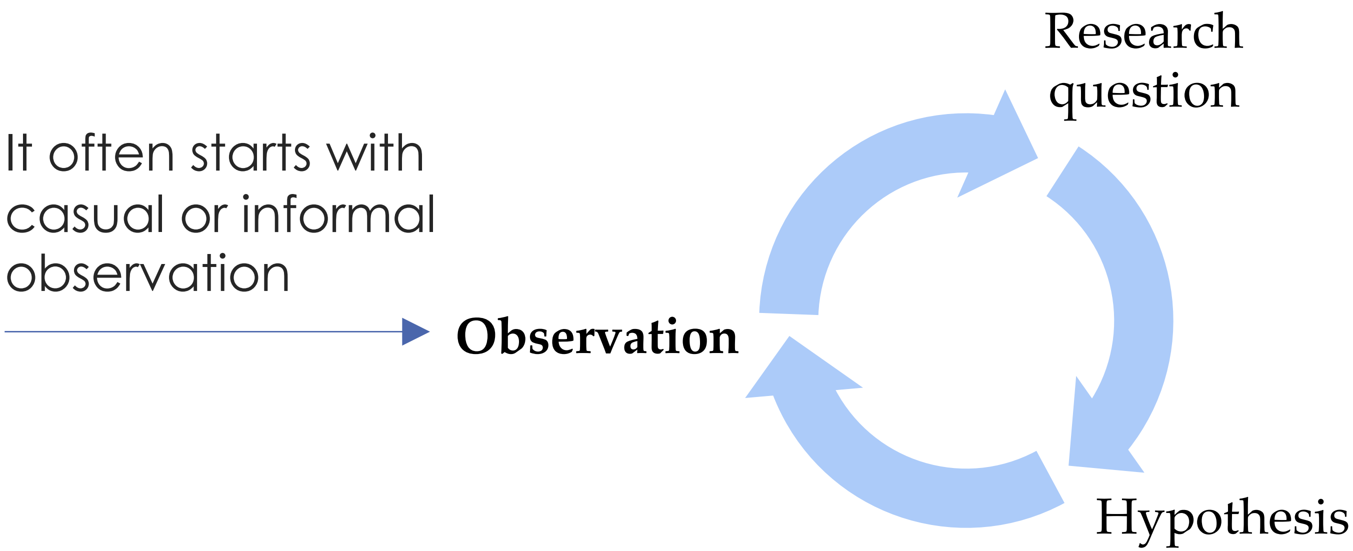
The ***empirical approach*** aims at making these observations "as general as possible", i.e., guaranteeing that we create valuable knowledge and not just data.

Empirical approach



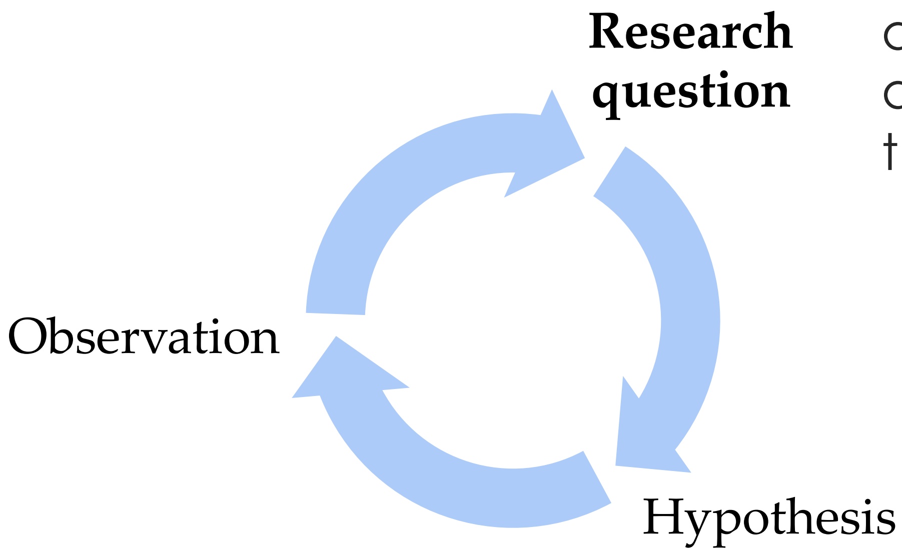
Empirical approach

It often starts with casual or informal observation



For example: Playing a mobile racing game via tilting is much harder than with touch buttons, but way more fun.

Empirical approach



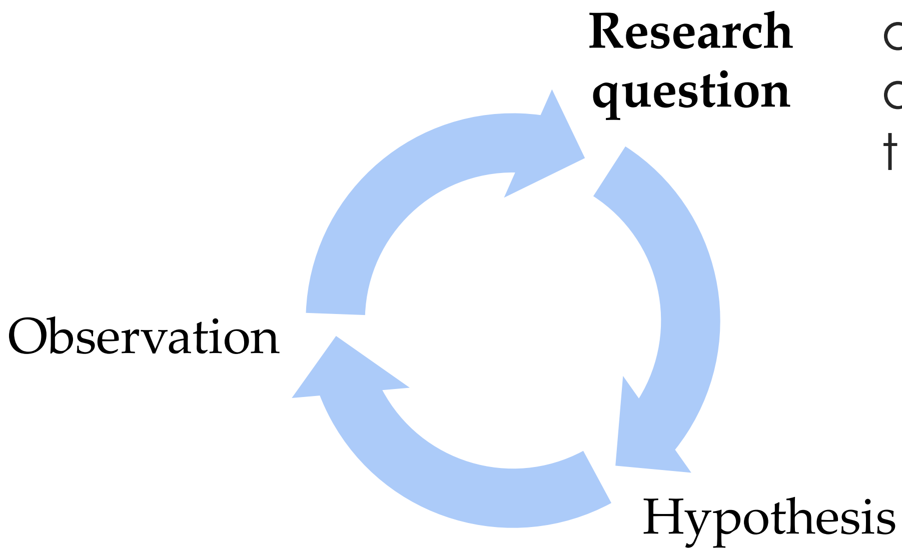
Identifies variables and hypothesis that are associated with the initial observation

Variables: characteristics or conditions that change or have different values for different individuals

Research question: a statement that describes or explains a relationship between or among variables

For example: How do performance and gameplay experience relate to interaction mode (tilt vs. touch) in mobile racing games?

Empirical approach



Identifies variables and hypothesis that are associated with the initial observation

Consider the following questions:

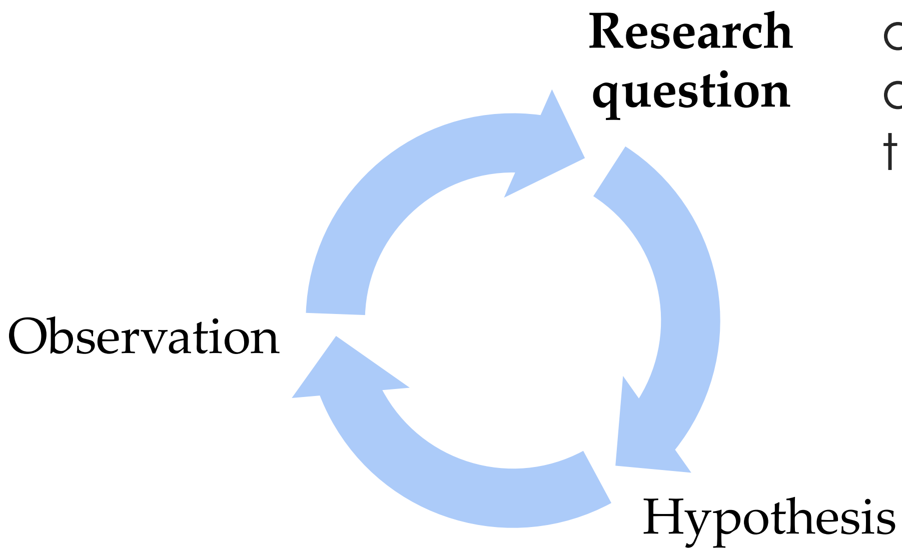
- Is it viable?
- Is it better than current practice?
- Which design alternative is best?
- What are the performance limits?
- What are the weaknesses?
- Does it work well for novices?
- How much practice is required?

⇒ These questions, while unquestionably relevant, are **not testable**.

⇒ Goal: transform these loose and informal questions to questions more suitable for empirical and experimental enquiry

Source: S. MacKenzie, CHI 2016 course on "Empirical Research Methods in HCI"

Empirical approach



Identifies variables and hypothesis that are associated with the initial observation

Exercise:

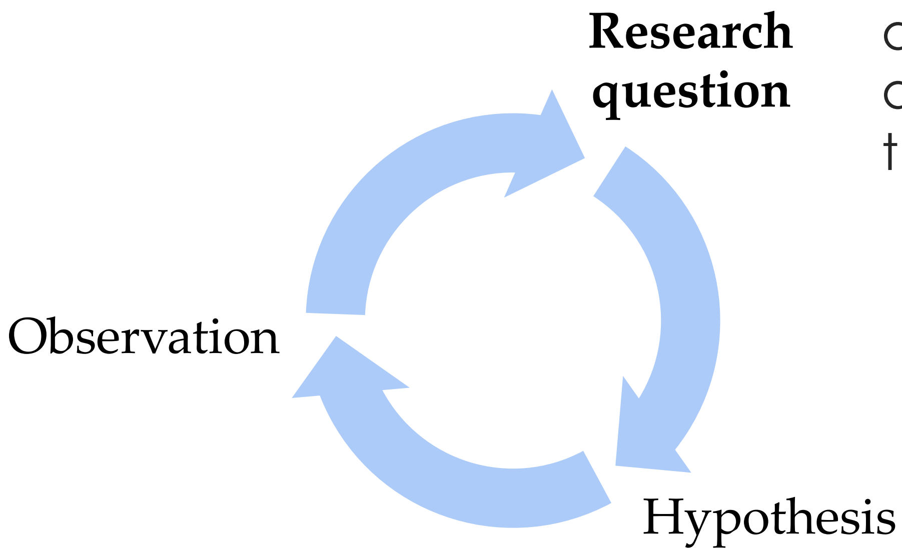
Try to re-cast as testable questions
(even though the new question may appear less important)

Scenario:

- You have invented a new text entry technique for mobile phones, and you think it is better than the existing Qwerty soft keyboard (QSK)
- You decide to undertake a program of empirical enquiry to evaluate your invention
- What are your research questions?

Source: S. MacKenzie, CHI 2016 course on "Empirical Research Methods in HCI"

Empirical approach



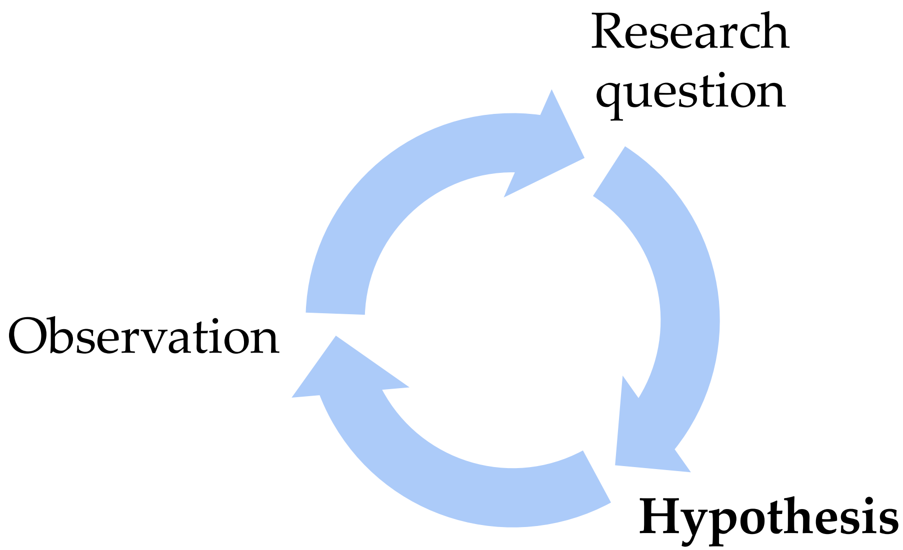
Identifies variables and hypothesis that are associated with the initial observation

Very weak:	Is the new technique any good?
Weak:	Is the new technique better than QSK?
Better:	Is the new technique faster than QSK?
Better still:	Is the new technique faster than QSK after a bit of practice?
Best:	Is the measured entry speed (in words per minute) higher for the new technique than for QSK after one hour of use?

Note how this narrows down the scope from a well-intended, broad yet untestable question to a narrower yet testable one.

Source: S. MacKenzie, CHI 2016 course on "Empirical Research Methods in HCI"

Empirical approach



Concrete and testable statements derived from the research question

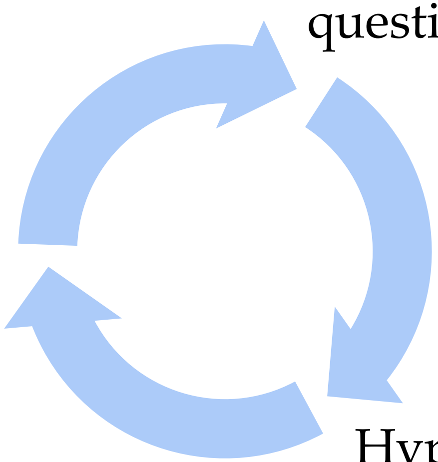
Generally, a specific set of operations for measuring external, observable behavior or changes. For example:

H1: Tilting has a negative effect on game performance high score compared to touch-based interaction in mobile racing games.

H2: Tilting has a positive effect on subjective game play experience compared to touch-based interaction in mobile racing games.

It usually predicts the outcome of the experiment.
The aim of the experiment is to proof or disproof this prediction.

Empirical approach



Collect data to support, refute, or refine the original hypothesis

Three strategies:

- **Descriptive research / observational method**: X happens
  Measures individual variable(s) to describe naturally-occurring phenomenon.
  E.g., rubber hand illusion (see first lecture).
- **Relational research / correlational method**: X and Y happen together
  Measures multiple variables for each participant
  E.g., is there a relation between performance and experience for tilting games (e.g., harder / more difficult = more fun).
- **Experimental research / experimental method**: X causes Y
  E.g., how does input mode (tilting vs. touch) influence performance for …?

Experimental research

Purpose: identify **cause-and-effect relationship**

- "Cause" is expressed by a controlled **independent variable** (aka **factor**), values of independent variables are sometimes called **levels.**
- "Effect" is measured by observed change of the **dependent variable(s)**

For example:

- *The independent variable "input mode" (tilt or touch) has an effect on the dependent variables "performance" (measured via high score) and "gameplay experience" (measured via standardized questionnaire results)*

Concrete factors and related levels are often referred to as **test conditions**.

Note that:

- Independent & dependent variables should follow directly from the research question.
- Having multiple independent variables is possible and common,
  but adds interaction effects (e.g., 2 variables: 3 effects, 3 variables: 7 effects, …)
  and increases the number of test conditions (critical when order is important)
- Other (external) influences might exist.
  They should be minimized or eliminated for controlled experiments.

Experimental research

Control variables = variables that might influence a dependent variable and are not under investigation, but can be controlled

Example: room lighting, temperature, background noise, …
Potential problem: can make results less generalizable.

Random variables = variables that are allowed to vary randomly

Introduce more variability and noise (bad),
but might make results more generalizable (good).

Confounding / extraneous variables = variables that systematically vary with the independent variable

Example 1: Three techniques A, B, and C are compared

- *All participants are tested on A, then on B, then on C => Performance may improve over time*
- *"Practice" is a confounding variable because it varies systematically with "technique"*

Example 2: Two search engine interfaces, Google and a new one, are compared

- *All participants have prior experience with Google, but none with the new one*
- *"Prior experience" is a confounding variable*

Note: In research papers you hardly ever find these terms,
but they are implicitly given when describing your experimental conditions.

Basic elements of an experimental study

For example: Input mode (tilt vs. touch)

Manipulation

Modify values of **independent variables**

For example: game score

Measurement

Measure value of **dependent variables** for each condition

For example: game score for tilt vs. touch

Comparison

Compare observations to support, refute, or refine hypothesis

Control

Control **other variables** to avoid influence on results

For example: age, gender, experience, location, …

Science is about creating new knowledge.
In the field of game & media technology, we often want to create new technology.
Human-centered research in GMT helps us achieve this, in two ways:

When developing a new technology (or concept, design, …),
research can answer questions such as:

- **Which abilities and limitations do humans possess that should be taken into account in the new design?**

Examples:

- What vibration patterns are recognizable by average humans? (And under what circumstances?)
- How many different vibration patterns can people remember?

When verifying if a new technology (or concept, design, …) works,
research can answer questions such as:

- **Does it meet its design goals?**

Examples:

- Does a new input device result in time savings/likeability?

In-class discussion: can you come up with an example related to mobile interaction (tilt versus touch) for games for the first type of research?

Empirical research is research using empirical evidence.
It is a way of gaining knowledge by means of direct and indirect observation or experience …

From Wikipedia ("Empirical research")

Empirical evidence is the information received by means of the senses, particularly by observation and documentation of patterns and behavior through experimentation.

From Wikipedia ("Empirical evidence")

Quote from a preceding slide about *experimental research*:
Answers to questions cannot be universal: whatever is observed is observed for the tested data only, and not for all conceivable data.

The ***empirical approach*** aims at making these observations "**as general as possible**", i.e., guaranteeing that we create valuable knowledge and not just data.

Internal and external validity

Note: External conditions *always* exist; influence on results needs to be minimized.

For example: If we cannot be sure that gender has no effect on the outcome, it becomes a random variable (external influence on results not controlled by test).

One way to deal with this is to run separate tests (i.e., make gender another independent variable) or to assign the same number of participants with the same gender to each tested condition (i.e., make it a control variable). Both require larger amounts of subjects!

They have an impact on **validity**:

- When controlled: higher confidence on causality → high **internal validity**
- When left as random: reflect variation in natural use → high **external validity**

Validity is a term describing the relevance and reliability of a result:

- **Internal validity**: to what degree can we assume that the effect does indeed result from the change of the independent variables?
- **External validity**: to what degree can we generalize our results to general conditions other than the ones under which we tested?

Tradeoff between internal & external validity

The more the test environment and experimental procedures are "relaxed" (to mimic –real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as pondering, distractions, or secondary tasks.

Internal & external validity are increased by posing multiple narrow (*testable*) questions that cover the range of outcomes influencing the broader (*untestable*) questions.

> E.g., a technique that is *faster*, is *more accurate*, takes *fewer steps*, is *easy to learn*, and is *easy to remember*, is generally *better*.

Fortunately there is usually a positive correlation between the *testable* and *untestable* questions.

> I.e., participants generally find a UI *better* if it is *faster, more accurate, takes fewer steps*, etc.

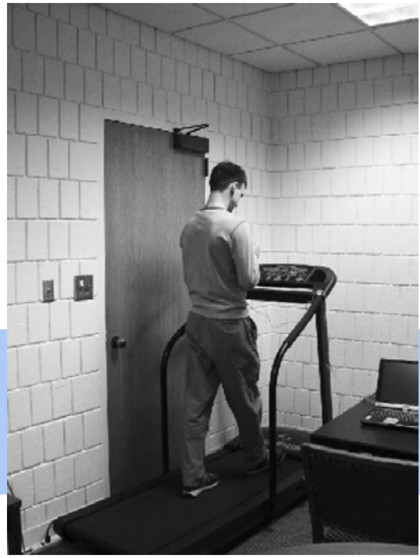
Again, remember the difference between the more general research aim or research problem and the very concrete research question (and how these should relate).

From: S. MacKenzie, CHI 2016 course on "Empirical Research Methods in HCI"

Where to do experimental research?

Controlled vs. "natural but uncontrolled" is particularly critical for **mobile** HCI!

Lab study



Real world study

(-) *artificial*
(+) *controlled*

(+) *realistic*
(-) *uncontrolled*

Environment and context matter,
even if they are not part of the actual experiment design!

Common setups in mobile HCI studies

Internal

Versus

External Validity

Lab study

+ fully controlled
- artificial, usually limited in time, size, subjects, …

Field study

+ more realistic (context matters!)
- less controlled, harder to interpret, complex, …

("Massive") online study

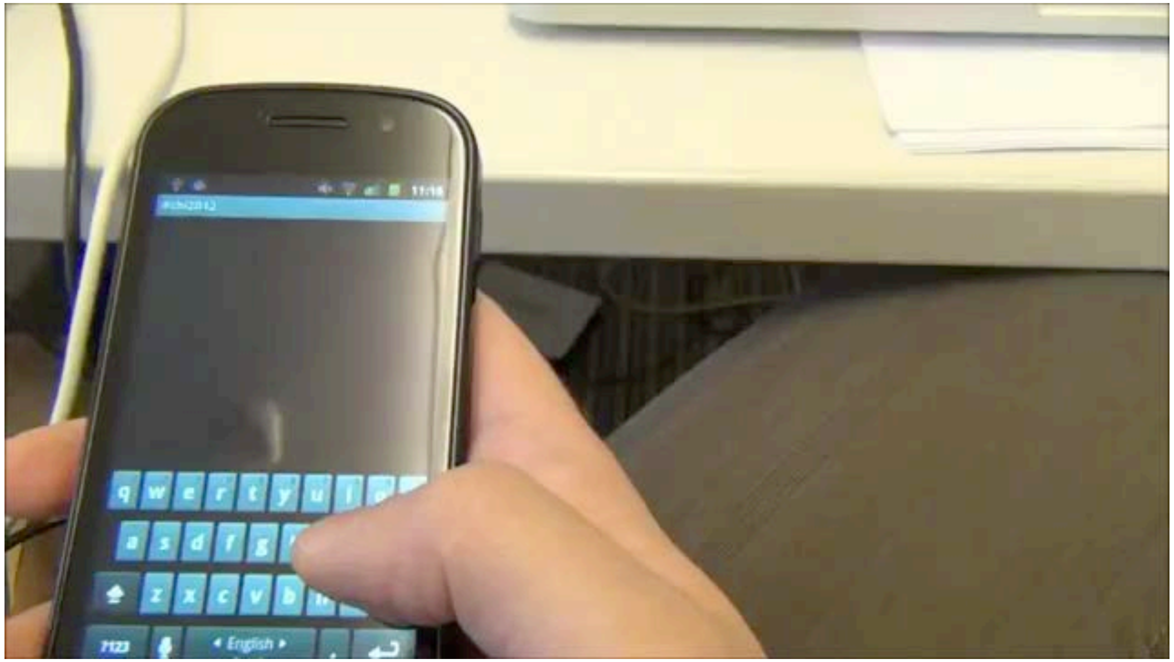
+ real usage (or not?)
- no control at all (and no "conditional knowledge")

Example for a (massive) online study

A large-scale study identifying a systematic skew in user's touch distribution on standard virtual keyboard

Comparison of three keyboard variations considering this observation:

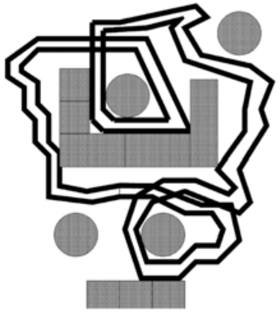
- Skew compensation
- Label shifting
- Touch position visualization



http://www.youtube.com/watch?v=DIfBsSLrvwU

Niels Henze, Enrico Rukzio, Susanne Boll: **Observational and Experimental Investigation of Typing Behaviour using Virtual Keyboards on Mobile Devices**, CHI 2012 (best paper award).

Comments on field studies



Important aspect in mobile: context, e.g., being in motion

- Guidelines from comparison of lab versus “artificial” walking conditions: treadmill can yield representative performance measures, but controlled walking scenarios more likely to adequately simulate actual user experience.
  Barnard et al. (2005) An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *Int. J. Hum.-Comput. Stud. 62, 4.*
- Significant differences found for lab versus field studies (e.g., frequency and severity of usability problems, user behavior & subjective responses.
  Duh et al. (2005) Usability evaluation for mobile device: a comparison of laboratory and field tests. *Proceedings of MobileHCI 2006.*
- Comparison of lab and **Experience Sampling Method (ESM)** showed that both can be informative to different aspects.
  Reyal et al. (2005) Performance and User Experience of Touchscreen and Gesture Keyboards in a Lab Setting and in the Wild. *Proceedings of ACM CHI 2015.*

Laboratory study

Users do the experiment in a dedicated place
selected and prepared by the experimenter

Potential advantages:

- Easier to use sophisticated equipment (A/V recording, two-way mirrors, …)
- Interruption free environment
- Full control (noise, lighting conditions, …)
- Safety (might be easier to maintain and guarantee) …
- …

Potential disadvantages:

- Lack of context, unnatural situation and environment (labs), …
- Not all environments can be simulated well enough
- …

Field study

Tests are done / observations & measurements take place
in the user's environment (or a real-world setting)

Potential advantages:

- Testing in a real (and realistic) context
- Can allow for longer studies (in some cases); e.g., even days or weeks

Potential disadvantages:

- External influences; e.g., high levels of noise, distractions, interruptions, …
- Safety can be an issue
- Test setup can still influence results (e.g., observer, equipment, …)

Note: sometimes borders are not strict
(e.g., "controlled" experiment in the real world)

Experimental research: study design

Correct **study / experiment design** is essential (and not easy):

- What variables to control (independent variables)?
- What variables to measure (dependent variables)?
- What test conditions, i.e., what levels (values or settings) to use for each independent variable?
- What tasks and procedures to use?

  Representative of actual usage (downside: more likely to include behaviors not directly related to the method under test)
- How many participants to use and how to solicit them?
- How to care for them, how to handle and protect them, how to follow ethical rules, …?
- Where to do it and how to control the environment (or not)?
- Other contexts that can, should, or shouldn't influence our results?
- Hardware and other equipment used for testing (and it's potential influence on results)

Etc.

Empirical research design is the major learning goal of assignment 5

Subjects / study participants

Who?

Should match expected target population or end user population (for user studies) as closely as possible.

Issues to consider include, but are not limited to:
similar age, level of education, experience (general and task domain)

Should be a sufficient number to allow for statistical testing and interpretation of the results.

Some advise:

- Use the same number of participants as in similar research (from respected sources).
- Also report how participants are selected (and be aware of drawbacks of *convenience sampling*).

Subjects / study participants

How many subjects? And who (male/female, age, experience, …)?
Hard to tell / no general rule (it depends on various factors)

Sample size must be large enough …

- to be representative for the population, …
- and take into account the design of the experiment …
- and the statistical methods chosen.

Note: sample size is not just number of users,
but also how often each condition (level per factor) is tested.

No golden rule or perfect answer exists;
it always depends on the situation and context.

Pragmatic considerations (availability of people, test equipment, time, …)
must be considered, but should not influence results.

Subjects / study participants

Who tests what? Which factors, which levels?
Again, this depends a lot on the context, situation, resources, …

Within-subjects design: each subject tests all options

(+) less subjects needed

(+) allows for qualitative comparison

(-) potential carryover or learning effects

⇒ Order of testing is very important / can be critical

Between-subjects design: each subject only tests one option

(+) shorter test duration

(+) no order effects (i.e., interference between conditions)

(-) variation not only within but also between subject groups

⇒ Assignment to groups is very important / can be critical (e.g., gender balance, age distribution, experience, …)

Mixed design: often used when multiple factors exist
(e.g., two-factor design: one within, one between)

Order effects

In a within-study design, the behavior may be influenced by experience that occurred earlier in the sequence[1].

Carryover effects: changes caused by the lingering aftereffects of an earlier treatment condition.

E.g., testing the first condition causes the finger to hurt, degrading the performance in the second condition

Progressive error: changes that are related to general experience in the study, but unrelated to specific treatment.

E.g., practice effects and fatigue (overall duration of experiment is too long)

Note: progressive errors can include learning effects, but they are not the same.

[1] Likewise, prior experience might impact results in both between- and within-subject designs.

Learning effects

Can happen in within- and between-study designs.

Learning curve: relationship between experience (or time) and performance.

Often: rapid raise at the beginning, followed by plateau.

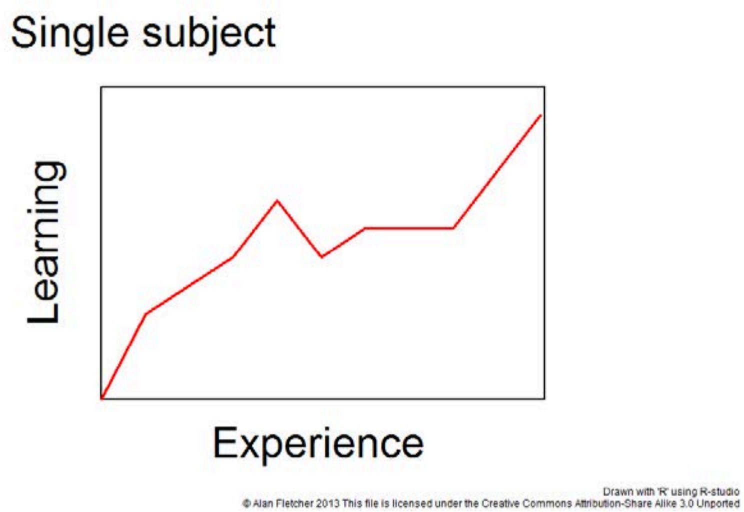
Measuring should start when learning effect is gone.

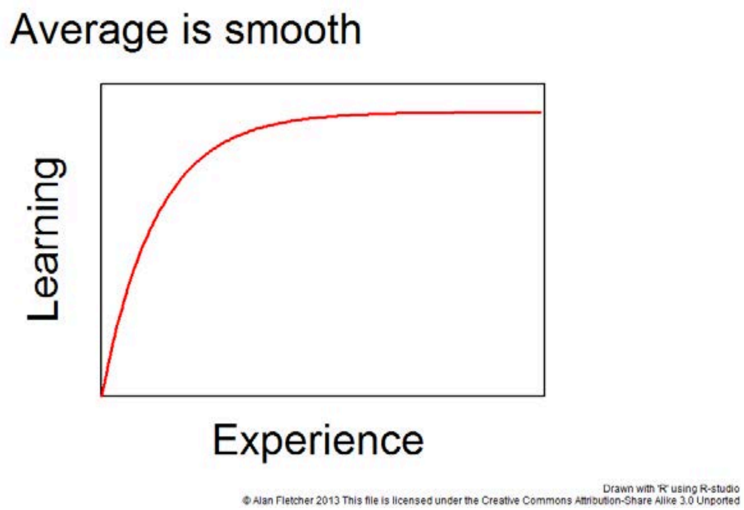
How?

- Tutorial?
- Practice task?
- Don't count first n samples?
  (Or measure when plateau is reached?)

A **learning curve** is a graphical representation of how an increase in learning (measured on the vertical axis) comes from greater experience (the horizontal axis); or how the more someone does something, the better they get at it.

From Wikipedia ("Learning curve")





Order in within-subject designs

How to avoid that participants in within-subject designs benefit from the first condition and consequently perform better in following ones?

Modify the **order of conditions**:

- Random
  in general only suitable for very large sample sizes
- Counterbalancing all possible orders
  make sure that it "fits" (e.g., 24 subjects and 4 options/conditions)
- Latin-square (each condition appears in each ordinal position, and each condition precedes and follows each condition one time)
  good if there are too many options

Counterbalancing vs. Latin square design

Example:

- *Assume you have **3 interfaces A, B, C** and want to split your participants in equally sized groups each doing the tests in a different order.*
- *How to map interface order to participant group?*

Counterbalanced mapping

	1	2	3
UI order 1	A	B	C
UI order 2	A	C	B
UI order 3	B	A	C
UI order 4	B	C	A
UI order 5	C	A	B
UI order 6	C	B	A

⇒ *Needs 6 equally sized groups of subjects*

Latin square

	1	2	3
UI order 1	A	B	C
UI order 2	C	A	B
UI order 3	B	C	A

Each option exactly once per row and once per column

⇒ *Needs 3 equally sized groups of subjects*

Question: can you identify a potential problem with this Latin square order?

Within subject-designs: potential pitfalls & issues

Don't forget ***confounding variables***!

E.g.: Three interfaces (e.g., video search UI),
tested with one test dataset (e.g., labeled video queries)

- Randomize order of video queries
- Split dataset in 3 and counterbalance it across interfaces
- …

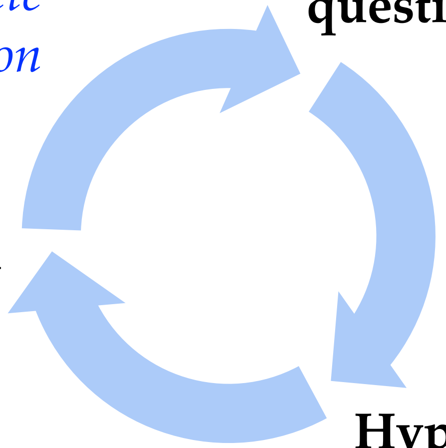
Just counterbalancing the order does ***not*** always eliminate order effects.

Putting subjects in groups of different order can make "order" another independent variable that needs to be analyzed.

Empirical approach

From research idea to concrete research question

Research question



From research question to concrete hypothesis

- *Variables: independent/dependent*
- *Type / strategy*

Hypothesis

Defines experiment design

Experiment design:

- *Subjects*
- *Environment / context*
- *Equipment (also for tests!)*

⇒ *Control, random, confounding variables*

Influences validity

Observation

Analysis of your data & results:

- *Statistical measures*
- *Interpretation*
- *Representation*

We will talk about this later.

Overview of (empirical) research

- Research goal / general aim
- Research question, hypothesis, other specification?
- Variables (independent, dependent)? Measures?
- Methodology?

- For empirical user studies: study design
  - Subjects: who, number, …
  - Environment
  - Equipment / material used
  - Other contexts
  - Tasks to perform, instructions given, …
  - Within/between-subjects design
  - Order (for within-subjects design) and other mappings (for both)

- Conclusions (answer to research question)
- Contributions (also with respect to goal/aim)

How about **ethics and integrity**?

Moral implications of my research?

Treatment of subjects and guaranteeing their wellbeing?

More on this later in relation to user studies.

Analysis of the results and conclusions drawn? (Correctness, flaws, honesty, …)

More on this later in relation to statistics.