Answering the right questions: Leveraging multiple methods to answer what you were looking for

Christine Bauer
Johannes Kepler Universität Linz
christine.bauer@jku.at
http://www.cp.jku.at
https://christinebauer.eu

Austrian Science Fund (FWF): V579.

JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Perception

ACM Summer School on Recommender Systems, *Gothenburg, Sweden, 13 September 2019*

Christine Bauer

Johannes Kepler University Linz, Austria
Institute of Computational Perception

Elise Richter excellence project, sponsored by Austrian Science Fund (FWF):
Fine-grained culture-aware music recommender systems

International Business Administration (Diploma)
Business Informatics (MSc)
Social and Economic Sciences (Business Informatics) (Doctorate)
Jazz Saxophone

In 1878 in Birka (Southeastern Sweden), unburied Viking settlement from about 750 to 950

High-status, Viking warrior, male.

Weapons found in the grave suggest the occupant was a high-status warrior. (Image credit: Neil Price, Charlotte Hedenstierna-Jonson, Torun Zachrisso, Anna Kjellström; Copyright : Antiquity Publications Ltd.)

Illustration how the burial might have looked just before it was closed in Viking times. (Image credit: Drawing by Þórhallur Þráinsson; Copyright Antiquity Publications Ltd.)

2017, DNA test

XX chromosomes

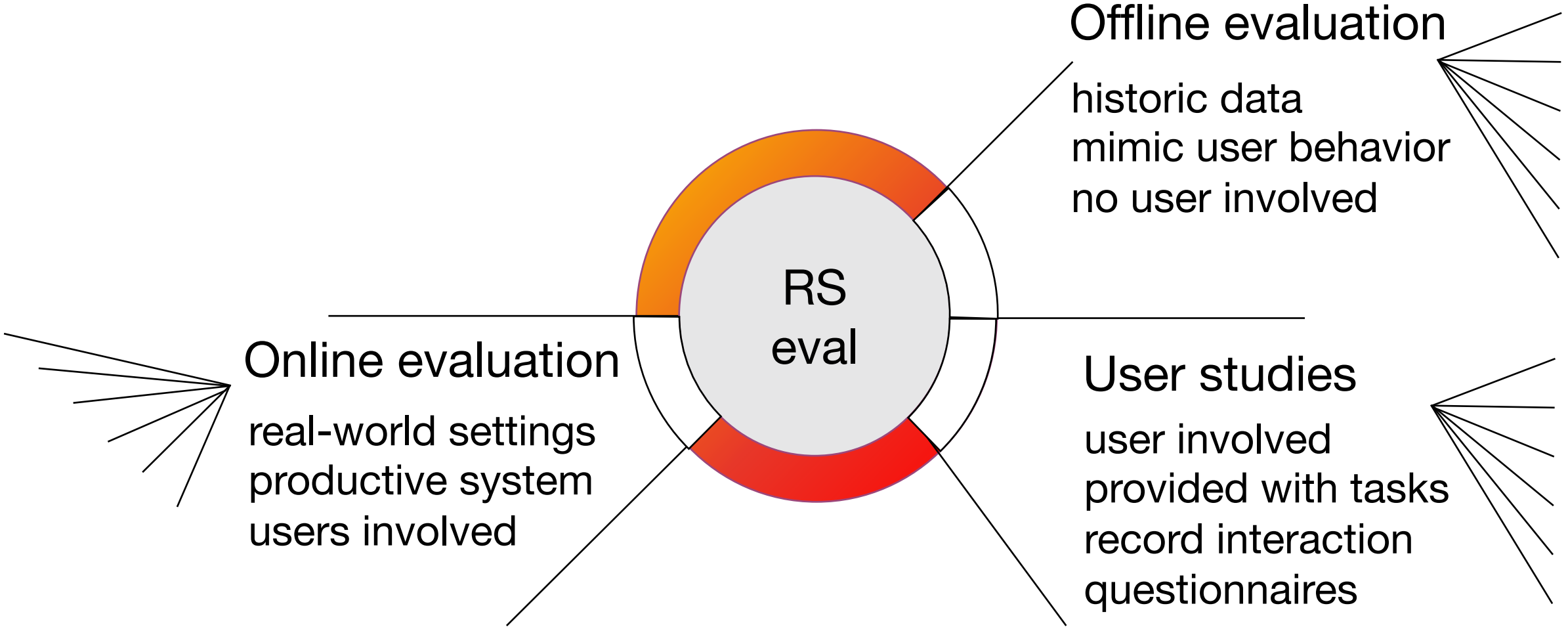
buried person is female

Assumption that arms, “non-female appearing” clothing in grave, etc. indicate a male warrior

Second method (other than looking on clothes and arms and setting) – the DNA test – gave a better insight

- All Viking warriors women?
- Was she a warrior?
- What was the social role of women in Viking time?
- ...

Tradition of recommender systems evaluation



Recall the example of choice overload from the Tuesday session

Less attractive
30% more sales
Higher purchase satisfaction

More attractive
3% more sales

Is the goal to increase sales?

Is the goal to have an attractive offer?

From Iyengar and Lepper (2000)

http://www.ted.com/talks/sheena_iyengar_choosing_what_to_choose.html (at 1:22)

There are blind spots in single method evaluation with one metric

Evaluating a music recommender system

Spotify
YouTube
pandora
...

example

Focus: Music consumer's perspective

Offline evaluation with focus on the music consumer

It can show that users' historic listening behavior can be simulated (e.g., high accuracy).

- Does the user want to listen to these familiar songs in the future? [redacted]
- Would the user be satisfied with the same number/proportion of unfamiliar songs? [redacted]
- Is the user interested in discovering (more) new, unfamiliar songs? [redacted]
- …

Online evaluation with focus on the music consumer

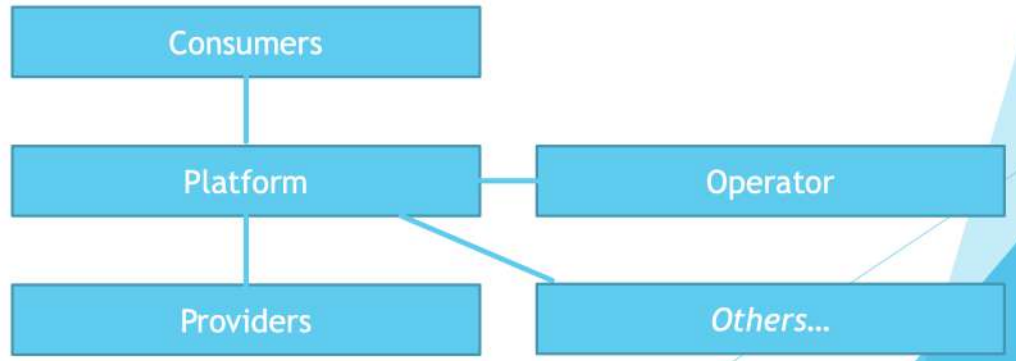
It can show that users click or skip recommended songs; or stay on platform for longer/shorter than usually.

- Does the user want to listen to the recommended songs in the future? [redacted]
- Is the user is satisfied with the number/proportion of unfamiliar songs recommended? [redacted]
  e.g., wants more discovery; skipped songs did not meet preferences; not in the mood for unfamiliar songs
- …

Recall yesterday's sessions on multi-stakeholder recommendations

Multisided platforms (MSPs)

- *Especially a system in which such parties lie on different sides of the recommendation interaction.*
- "Multisided platforms are technologies, products or services that create value primarily by enabling direct interactions between two or more customer or participant groups." (Hagiu, 2014)



Robin Burke (2019)

There are various stakeholder involved in a music recommender ecosystem

- music consumer
- the society
- platform provider
- music company
- top-of-the-top-superstar
- “long tail” artist

Spotify

YouTube

pandora

UNIVERSAL

WARNER MUSIC GROUP

SONY MUSIC

Routenote

SOUNDROP

What does all that mean for evaluation?

Results of the tradition of recommender systems evaluation

Focus on one single perspective

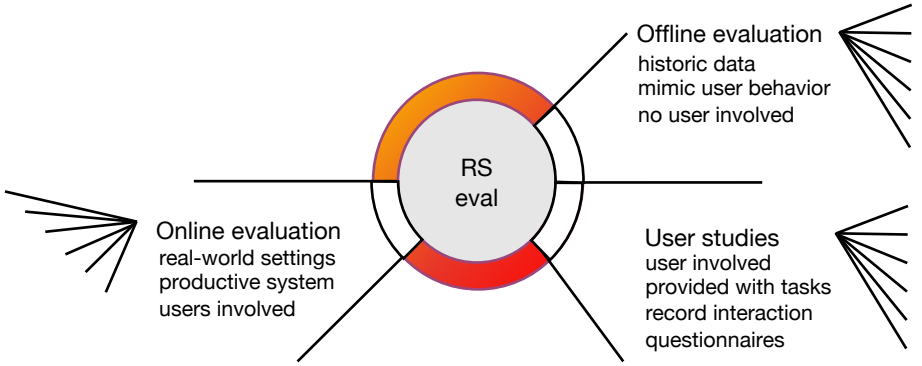
Incomplete picture: blind spots

Small set of metrics;
often picked from one perspective only

Evaluation results may differ

e.g., user satisfaction does not always correlate with high recommender accuracy

offline evaluations of accuracy are not always meaningful for predicting relative performance of different techniques



Multi-method evaluation

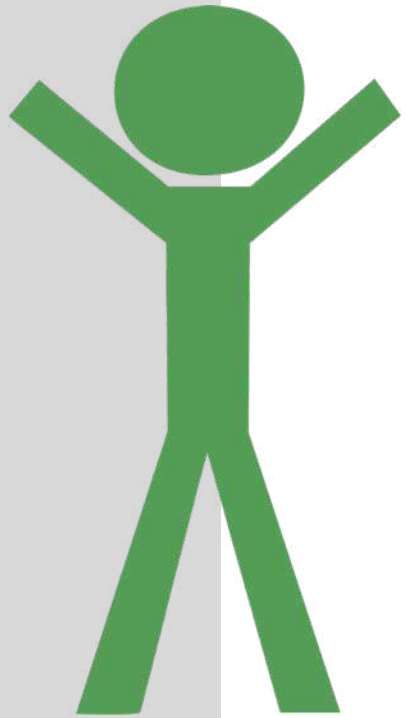
Goal:
Getting an integrated big picture of recommender system performance

Combine several (quantitative and/or qualitative) evaluation methods

- To capture the same phenomenon from different angles
- To capture diverse, but complementary phenomena
- To resolve conflicting findings
- To get an integrated picture of performance in the context of use
- To triangulate quality

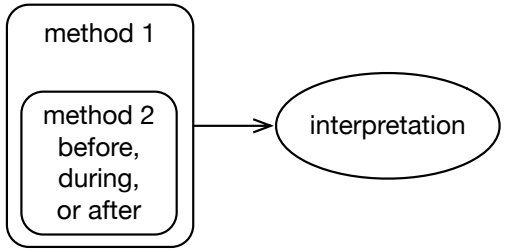
Benefits

- Explore sophisticated issues more holistically and widely
- Capture diverse, but complementary phenomena
- Apply diverse methods to capture the same phenomenon from possibly different angles
- Resolve conflicting findings
- Neutralize biases inherent to evaluation approaches

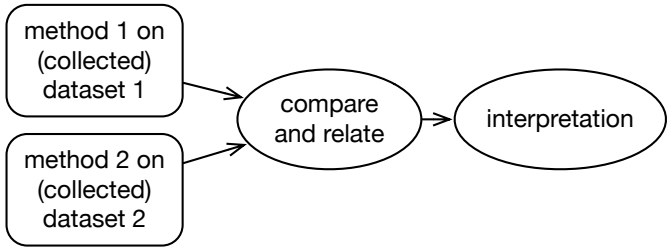


There are several strategies for multi-method evaluation

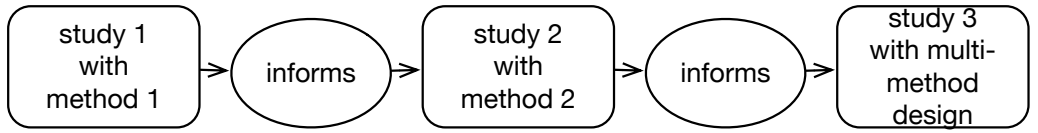
John W Creswell and Vicki L. Plano Clark. 2011. Designing and conducting mixed methods research. Sage Publications, Los Angeles, CA, USA.



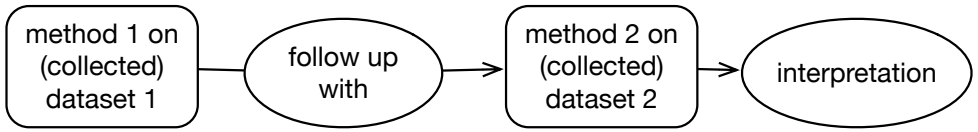
(c) The embedded design



(a) The convergent parallel design

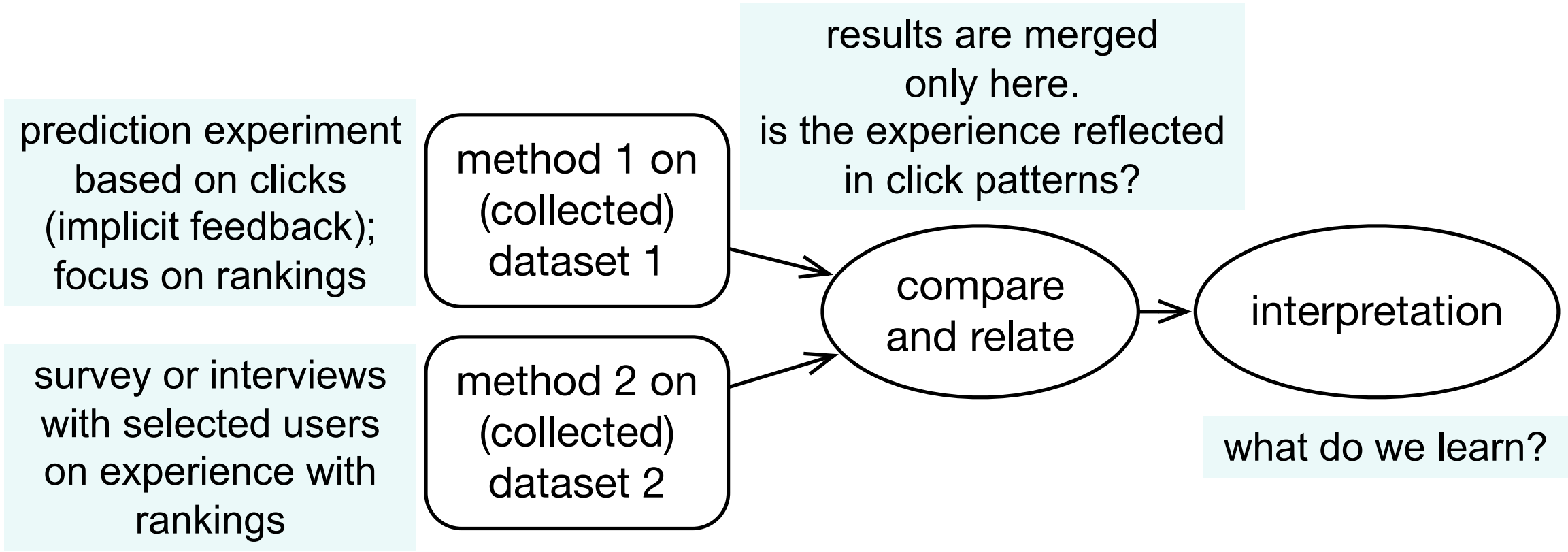


(d) The multi-phase design



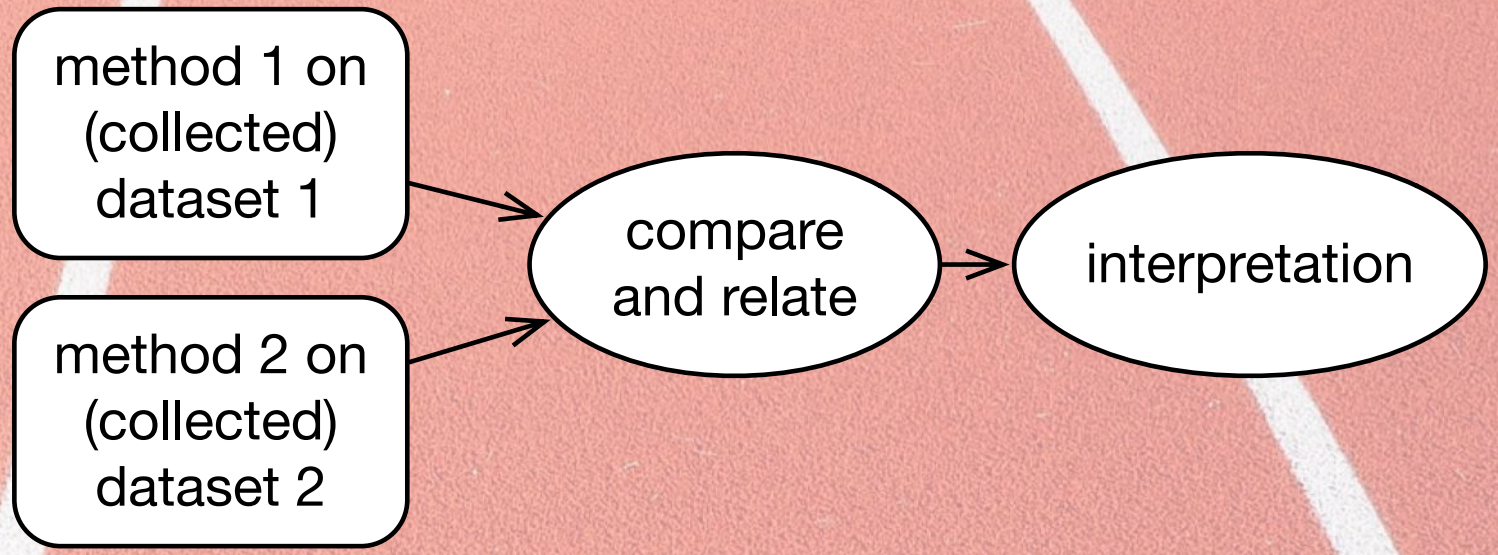
(b) The sequential design

(a) The convergent parallel design

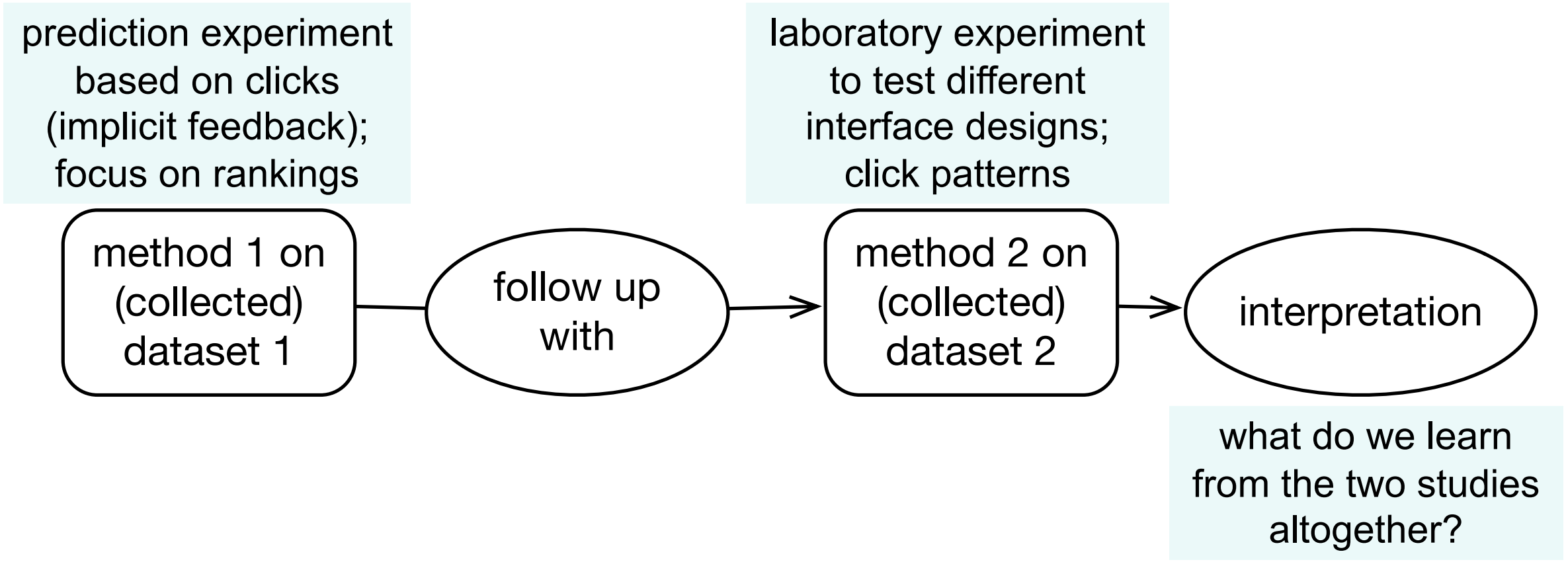


Your turn!

- Find a sparring partner in the room
- Take your research goal and question
- How can you address it in a convergent parallel design?
- Discuss!

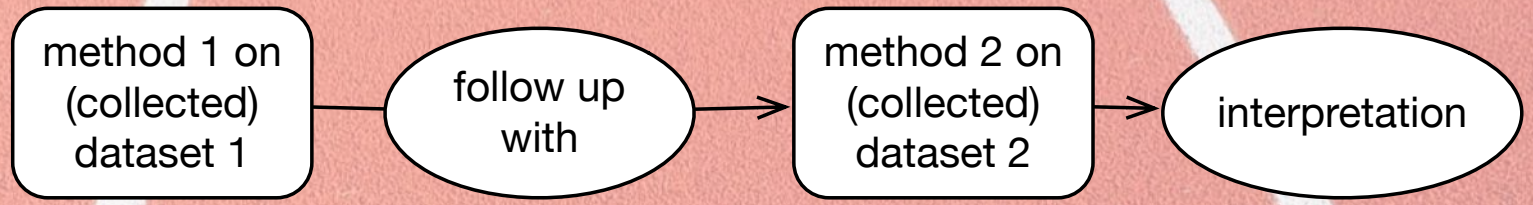


(b) The sequential design



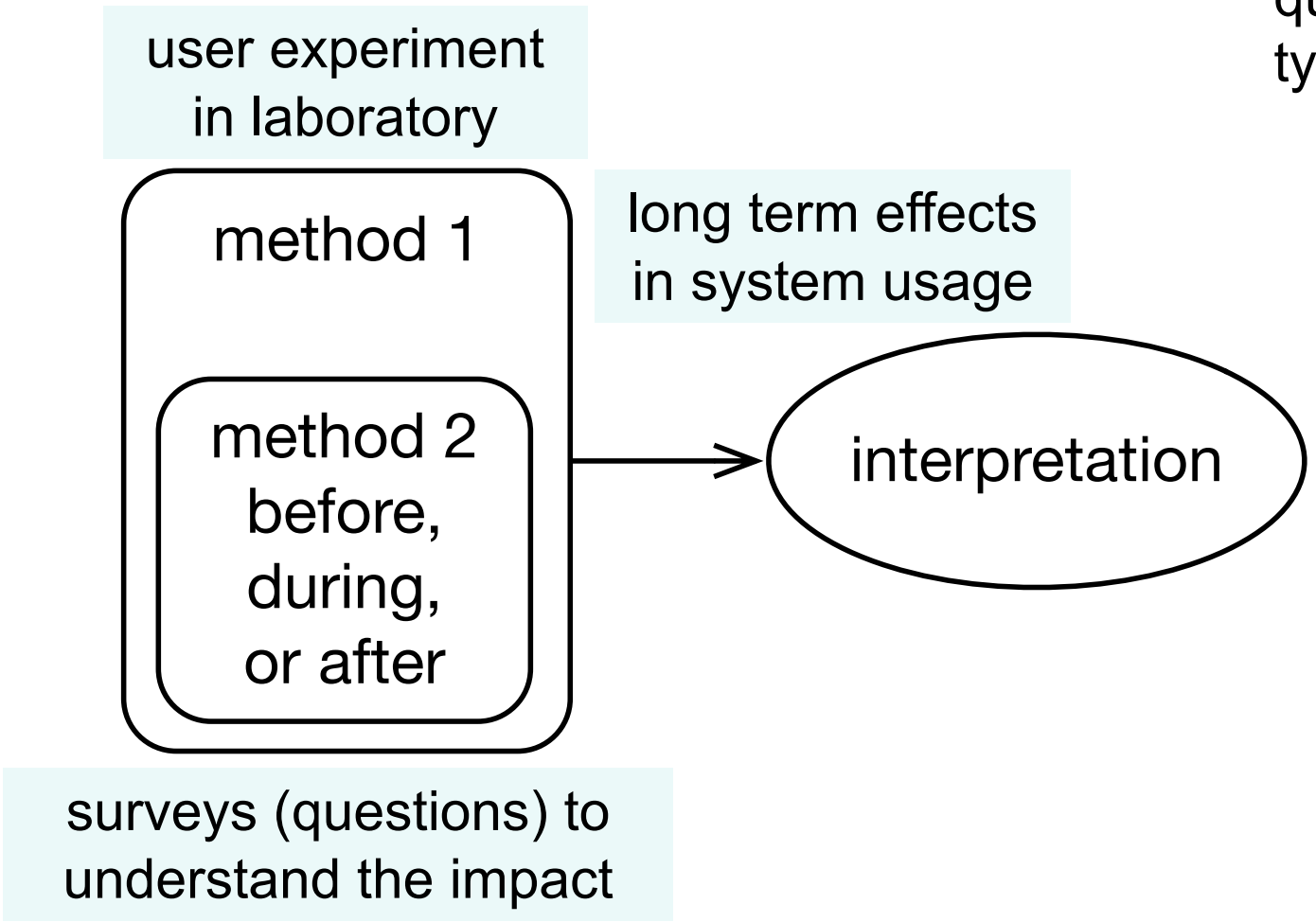
Your turn!

- Find a sparring partner in the room
- Take your research goal and question
- How can you address it in a sequential design?
- Discuss!



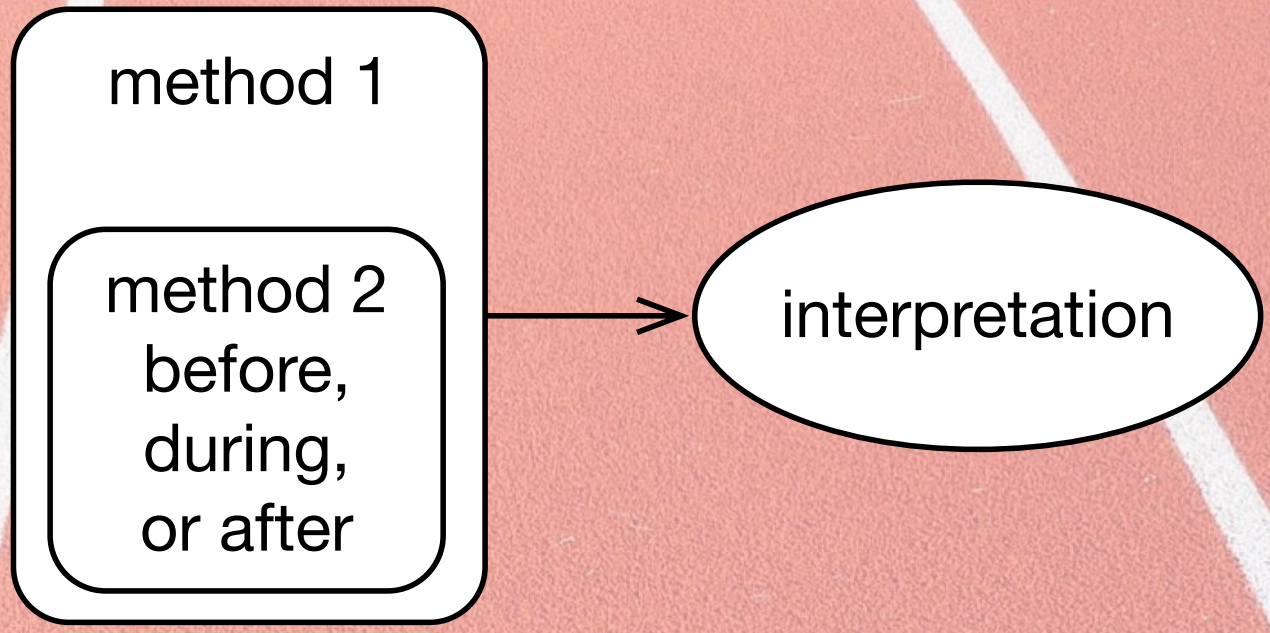
(c) The embedded design

Purpose is to answer different questions that require different types of data.



Your turn!

- Find a sparring partner in the room
- Take your research goal and question
- How can you address it in an embedded design?
- Discuss!

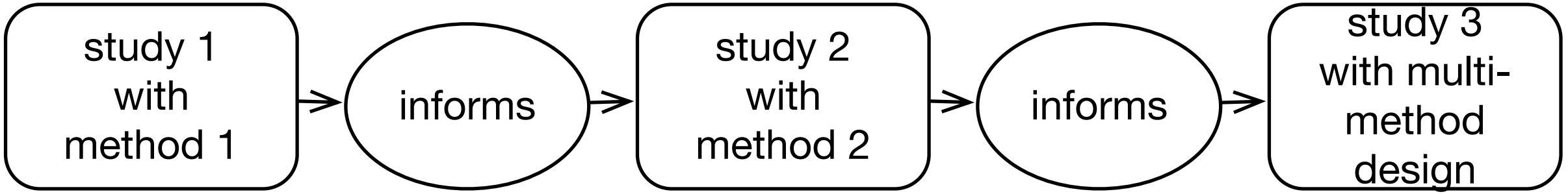


(d) The multi-phase design

online study focusing on click patterns of users using a recsys

think-aloud study with selected users with goal to find out why they click on which item or quit

experiment to test influencing factors on different clicking behavior with additional survey



we find out that users click mostly on the first three items, then they quit the platform

what do we learn from this altogether?

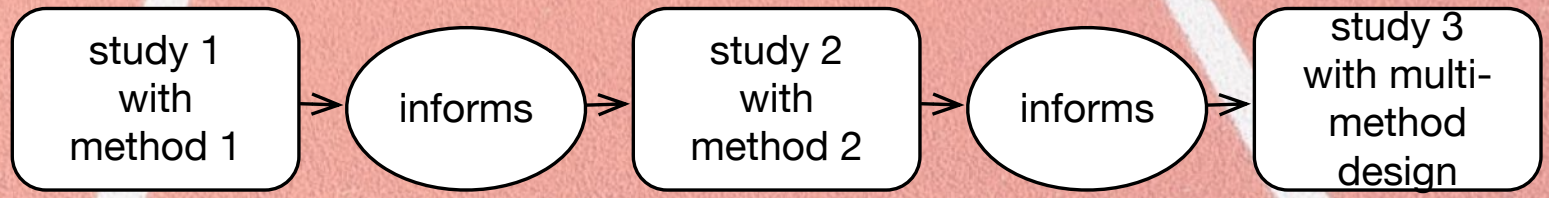
Your turn!

Find a sparring partner in the room

Take your research goal and question

How can you address it in a multi-phase design?

Discuss!



Your turn!

Find a sparring partner in the room

Take your research goal and question

Do all four versions equally make sense for your research goal? Why? Why not?

Discuss!

Things to remember

- Look at phenomena from different angles
- If your research is related to users, involve them!
- When focusing, have the overall picture in mind
- When having the overall picture in mind, keep your focus

Sounds easy ☺

However,
what about the details?

Where should we start?

Every research endeavor starts with a **research goal** and a **research question**.

What is a research question?

The researcher asks a very specific question and tests a specific hypothesis.

Broad questions are usually broken into smaller, testable hypotheses or questions.

Often called an objective or aim, though calling it a question tends to help with focusing and thinking about how to find an answer.

What makes a poor research question?

a question that matters to nobody, not even you

hoping one emerges from routine records/data available

- the records will be biased and confounded
- they will lack information you need to answer your question reliably, because they were collected for another reason

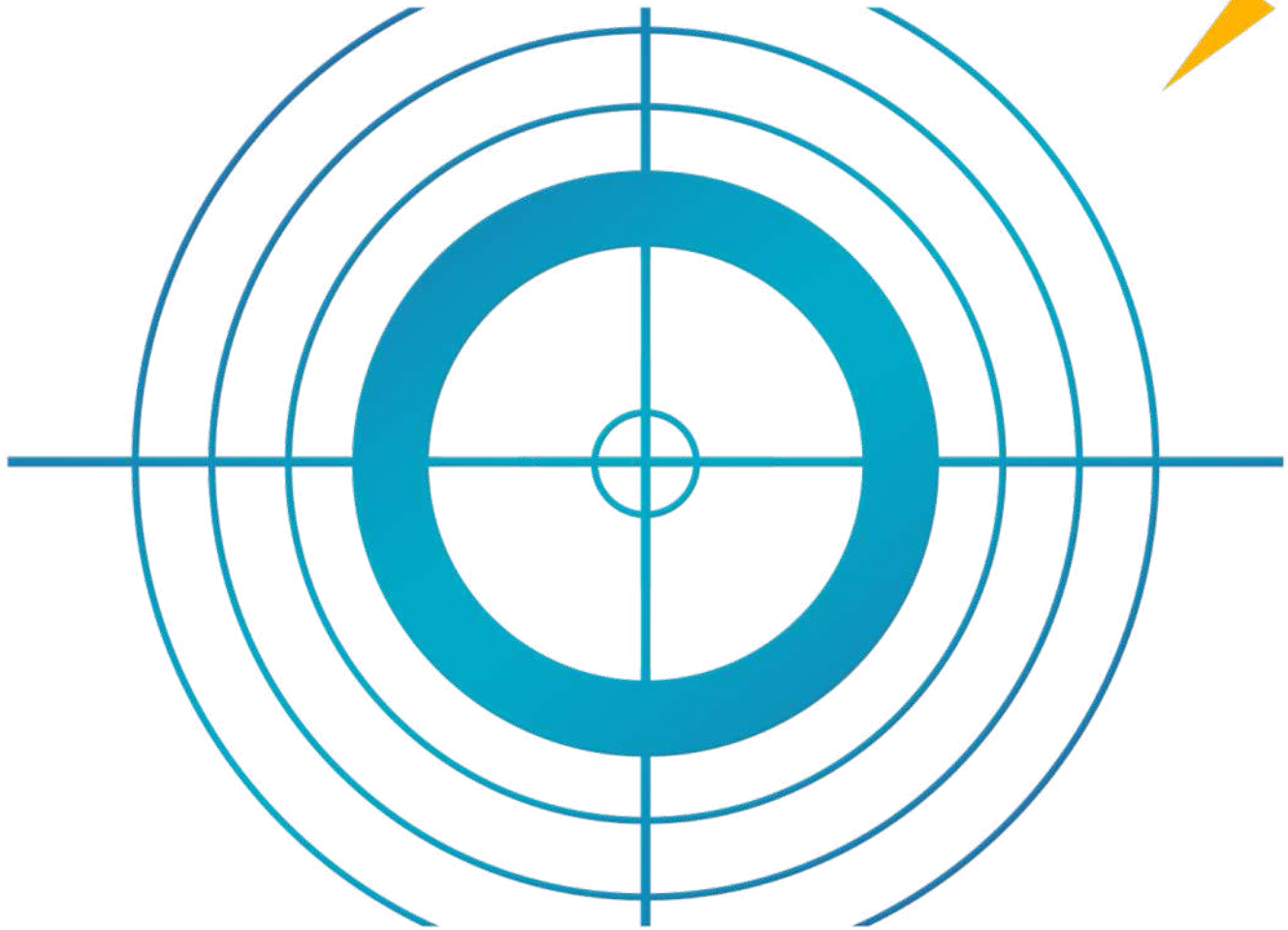
fishing expedition/data dredging – gathering new data and hoping a question will emerge

What makes a good question?

Specificity

Focus

!!



PICO

Problem or Population

- Who is the population (target subjects) or what's the problem?

Intervention

- What is the intervention or exposure?

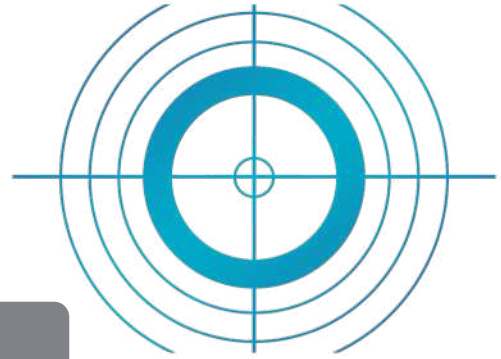
Comparison

- What is the comparison group/algorithm/setting/...?

Outcome

- What is the outcome or endpoint of the study?

How to focus your question?



Some ideas…

- (brief) literature search for previous evidence
- discuss with colleagues (e.g., peers, experts, non-experts,…)
- narrow down the question (e.g., time, place, group)
- what answer do you expect to find?

Your turn!

Form groups of ~3-4

Take your research goal and question and go through PICO

Discuss with your peers

Problem or Population

- Who is the population (target subjects) or what's the problem?

Intervention

- What is the intervention or exposure?

Comparison

- What is the comparison group/algorithm/setting/...?

Outcome

- What is the outcome or endpoint of the study?

Now what?

From a research question to a **study design**

Importance of study design

It will determine how you collect, analyze and interpret your data.

It helps you decide what resources you need.

It has an impact on the reliability of your study results.

Types of study designs

Descriptive

- Provides an overview of what is happening within a particular population or group
- Includes – but is not restricted to – qualitative

Analytical

- Quantify the relationship between two factors
  - Experimental designs
  - Observational designs

Guiding questions

- What is the research question?
- Who am I collecting information from?
- What kinds of information do I need?
- How much information will I need?
  - sample size – ask a statistician for help
  - information points per subject in sample
- How will I use the information?
- How will I minimize chance/bias/confounding?
- What is ethically appropriate?
  - How will I collect the information ethically?
- What is feasible?

Your turn!

Form groups of ~3-4

Take your research goal and question and go through these questions

Discuss with your peers

- What is the research question?
- Who am I collecting information from?
- What kinds of information do I need?
- How much information will I need?
- How will I use the information?
- How will I minimize chance/bias/confounding?
- What is ethically appropriate?
- What is feasible?

Factors to consider when choosing one method over another?

Balance between strengths and weaknesses associated with each method

Time for data collection and analysis

- observation or interview method helps to collect richer information, but it takes time
- survey helps you collect more data quickly, yet it may lack details

Feasibility of data acquisition / access to data

- dataset available that really fits the research goal
  (e.g., MovieLens again? Yes/no? Why/why not?)
- access to target group
  (access to specific user groups may be challenging; e.g., children, experts in a field)
- privacy and ethical concerns (institutional review board (IRB))

Access to skills for the method

- being non-skilled is not an excuse!!
- learning takes time
- identifying and getting involved skilled co-contributors takes time

...

What is feasible?
Pilot studies

Small scale preliminary study of your larger trial

Helps to establish…

- Feasibility
- Procedures and materials
- Cost
- Barriers and enablers
- Track record and team cohesion

How will you reduce bias?

Strategies to reduce error within your data.

How to avoid confounding.

Reduce threats to the validity of your study.

Things to remember

- Select a design that allows you to answer your research question
- Select a design that provides the highest level of evidence possible – but is also feasible
- Conduct a pilot
- Pay attention to the finer details

Checklist for the research project (1/3)
Before you believe that your are done, check again!

1. Theoretical feasibility

- You can't do it all by yourself…
- Check the literature (overview and "ground breaking" articles are particularly helpful)
- Speak to advisors, peers, and other researchers in the field

2. Inventory of approaches and methods

- There exist constraints. You can't study everything in any way, but you do have choice
- Finding a good research design always is an iterative process (so don't worry if your first version looks bad)

3. Identify your main approach

- Look at similar research
- Remember, if you want to "prove things" and make causality claims, you need comparison!
- e.g., qualitative approaches to explore and to understand, quantitative approaches to confirm, generalize, prove,...

Checklist for the research project (1/3)
Before you believe that your are done, check again!

4. Methodological feasibility

- Make a list of all the concepts that appear in your research questions (and hypothesis if you have)
- Take each concept apart for its dimensions
- Operationalize each empirical dimension (make it is measurable)

5. Does your theory part really relate to your empirical / practical part?

Checklist for the research project (1/3)
Before you believe that your are done, check again!

6. Make sure that you can produce needed data and then analyze them

- Do you know how to gather data (make observations, design questionnaires, make interviews,...)

7. Check your skills and resources

- Can you handle these data?
- Can someone assist?

8. Do some planning

Take home message

- Research design is a function of the research question; *not* choice!
- Match the research design to research question
- Focus, focus, focus!

Answering the right questions: Leveraging multiple methods to answer what you were looking for

Christine Bauer
Johannes Kepler Universität Linz
christine.bauer@jku.at
http://www.cp.jku.at
https://christinebauer.eu

Austrian Science Fund (FWF): V579.

JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Perception

ACM Summer School on Recommender Systems, *Gothenburg, Sweden, 13 September 2019*