# 

## Research Questions

Research Methods for Human-Centered Computing



Today's goal:

Teach you how to develop research questions for your study proposal

Outline:

- What are research questions?
- Where do they come from?
- What makes a research questions good?
- Brainstorming research questions



#### Please do the pre-test on Canvas!



# **Definition**

What is a research question?



A research question defines the objective of a study

Benefits for you, before doing the study:

- It distills your research topic area into a specific goal
- It motivates the design of your study
- It drives the development of testable hypotheses

Benefits for you, after the study:

- It motivates the analysis of collected data
- It scopes the argumentation of your paper



A research question defines the objective of a study

Benefits for others, before the study:

It explains what question you set out to answer in a study proposal (to IRB, grant reviewers, potential collaborators)

Benefits for others, after the study:

It explains to readers what they will learn from your paper It explains to listeners what they will learn from your presentation



Collective Privacy Management

The management of information with multiple stakeholders

Examples:

- Tagged photos
- Posts by others that may reveal information about me
- Posts by me that others may not want to see



Many social dynamics on social media are culturally embedded:

- Social capital formation
- Relationship maintenance

What if collective privacy management is also a culturally shaped phenomenon?



#### **Research questions:**

Do these practices have different consequences in different countries, in terms of amount, accuracy, and valence of information sharing?

Are these practices conceptualized differently between countries, which means that they should also be measured differently?



amount, accuracy, valence (see paper)





Research questions:

Questions

Usually open-ended

General (each paper usually has 1–3 research questions)

Describes broad relationships

#### Hypotheses:

Statements

Supported or rejected

Specific (each research question can lead to multiple hypotheses)

Describes relationships between specific variables



#### Research question:

How come personalized video clip recommendations **reduce** users' overall viewing behavior (compared to random clips)?



Hypotheses:

- Users perceive personalized recommendations to have a higher quality than random recommendations
- Recommendation quality is positively associated with perceived system effectiveness
- Recommendation quality is positively associated with choice satisfaction
- Perceived system effectiveness is **negatively** related to users' total viewing time
- Perceived system effectiveness is **negatively** related to the number of clips people watch







## Source

Where do research questions come from?



#### Research topic area



#### Specific research question(s)



Topic areas can come from anywhere!

- Your curiosity
- Your advisor's grant topic
- "Hot" topics
- Really anywhere!
  - Seminars, conversations with others, newspapers, podcasts



Types of topic areas:

- A (feature of) technology that fulfills a certain human or societal need
- A situation where technology creates a new social dynamic or a social/societal problem
- A new way of interacting with the digital world
- A situation that could benefit from (a fundamental rethinking of) digital assistance



A (feature of) technology that fulfills a certain personal or societal need:

- Recommender systems for privacy
- Diversity in videogames

A situation where technology creates a new social dynamic or a social/societal problem:

- The effect of Facebook on non-users
- Defaults and framing lead to worse privacy decisionmaking



A new way of interacting with the digital world:

- The effects of virtual agents
- "Glanceable" interfaces

A situation that could benefit from (a fundamental rethinking of) digital assistance:

- Social event planning
- Helping older folks to remember to take their medications



They can come from:

- Again, your curiosity
- Practical problems

- Evaluate an intervention, assumption or new technology

- Requests for proposals
- Literature
  - Previous studies
  - Theory



Advanced Distributed Ceaning

#### **BAA Update**

#### **Topic: Activity Registry and Common Course Catalog**

Across the Department of Defense (DoD), training and education systems historically use pre-determined, point-to-point connections and manual intervention to share actionable information relating to requirements, curricula, and student performance across systems. This requires integrators to customize each system to get the data out of one and into another, and human processes to analyzes and correlate it. This is especially true with the way we catalog and archive different types of instructional activities including digital content, instructor-led courses, simulations, and other learning opportunities. This is complicated in that a single entity does not own all training and education content and therefore can only publish catalogs (e.g. CANTRAC) for the courses they own.





In the face of numerous, varied, and complex information sharing challenges facing the DoD, the DoD Chief Information Officer (CIO) has set a vision to deliver an Information Enterprise (IE) that enables DoD and partners to securely access data and information services they need at the right time and place, and on approved devices of their choosing. The DoD Net-Centric Data Strategy provides the basis for implementing and sharing data in a net-centric environment. It describes the requirements for inputting and sharing data, but it does not accommodate the myriad of complexities associated with today's training and education environment. The Total Learning Architecture (TLA) data strategy summarized in Figure 1, along with the related common course catalog from the DoD Chief Management Officer, build upon the DoD Net-Centric Data Strategy's goals of making data assets visible, accessible, secure, and understandable. This approach establishes services as the preferred means by which information producers and capability providers can make their data assets and

associated operational activities available across DoD and beyond. It also establishes services (i.e. data contracts and manipulation business logic) as the preferred means by which consumers can access and use these data assets and capabilities.

# Proposal requests

As envisioned, this research will focus on identifying, enhancing, or developing the standards and specifications necessary to create a federated data framework that allows individual DoD components to manage their own learning resources via an Activity Registry that rolls up (i.e. federates) into a Common Course Catalog. The key attributes of this research include:

- Ensure data about learning resources is visible, secure, available, and usable when needed and where needed to accelerate decision-making. Predict and inventory requirements for learning activities (e.g., courses) to meet the demands of DoD consumers while minimizing the duplication of course development efforts.
- Align with the DoD Discovery Metadata Specification (DDMS) which specifies how to use metadata to make data assets visible across the enterprise.
- Automate the ability to align content with competencies, generate and manage paradata, perform semantic search services, and create machine-actionable metadata that roll up from local Activity Registries into a single Common Course Catalog.
- Align ongoing initiatives from the DoD CIO, CMO, and the Chief Data Officer with the needs of the training and education community, while providing guidance to ongoing training and education modernization efforts across DoD.

The ADL Initiative invites white paper proposals to address these issues through the following tasks:

- Task 1: Definition/Validation of the learning activity metadata elements including content, courses, exercises, and other formal learning activities common across the DoD. Identify an approach for describing various learning activities a learner will encounter across the continuum of lifelong learning. Investigate various metadata formats, standards and specifications to inform an approach that describes content in support of the Future Learning Ecosystem. Create an approach that considers machine-learning created metadata (for example, relevance and engagement) beyond the more commonly available human-specified attributes. This task includes consideration of metadata storage and retrieval methods across the myriad of T&E commands within DoD (e.g., Activity Indices). The deliverables would include both Metadata Strategy and Metadata Guidance.
- Task 2: Definition/Validation of learning activity paradata elements (data about usage). Identify a taxonomy and approach for gathering, verifying, storing, and sharing paradata within an Activity Registry. While metadata classifies properties about the learning resource, paradata involves the collection and open sharing of in situ information about users' actions, preferences or experiences related to the resource. Paradata is generated through user processes of searching for





#### About The Program

Eligibility Guidelines

The NSF Graduate Research Fellowship Program recognizes and supports outstanding graduate students in NSF-supported science, technology, engineering, and mathematics disciplines who are pursuing researchbased Master's and doctoral segrees at accredited United States institutions.

Read More +

See the current Program Solicitation for eligibility guidelines. Those already enrolled in graduate school may apply one time only.

#### In The News

#### New NSF Funding Opportunity:

INTERN (Non-Asademia Research Internships for Graduate Students Supplemental Funding)

NSF has identified improvement of graduate student preparedness for the Science, Technology, Engineering and Mathematics (STEM) workforce as one of its priorities. As part of this effort, a supplemental funding opportunity is available in fiscal year (FY) 2018 and FY 2019. This lunding provides support for non-academic research internships for graduate students for career opportunities in any sector of the U.S. economy.



Initial work

- Pilot studies, case studies, studies with a small N
- Questionable assumptions
  - E.g.: "More sharing options is better for privacy"
- Conflicting findings
  - E.g.: Facebook use and depression
- Overlooked variables
  - E.g.: The moderating effect of culture



#### John, Dave, Ethan



nothing city city block exact location Outside work hrs nothing city city block exact location

## Other friends

#### Ben, Robert, Thomas, Mike, Paul, Stev Frank

During work hrs:

Outside work hrs



\_

. . .



Setting and expanding boundaries

E.g.: Does choice overload happen in recommender systems?

**Testing alternative explanations (theories)** See next slide



#### John, Dave, Ethan

During work hrs:



Outside work hrs
nothing

city city block exact location

## Other friends

#### Ben, Robert, Thomas, Mike, Paul, Stev Frank

During work hrs:

Outside work hrs



\_

. . .



Purposes of reviewing the literature:

- Provide a context for your research
- Avoid duplication effort
- Argue relevance of your work
- Find relevant theories
- Identify potential problems in conducting the research
- Identify "acceptable" practices of the field

Resources:

dl.acm.org, ieeexplore.ieee.org, scholar.google.com



Unlike other fields, HCI does not restrict you to a specific set of journals/conferences

In fact, reviewers will usually demand that you look beyond the core field of HCI!

What are good and bad sources?



Credible, peer-reviewed journals

Large publisher, "famous" editorial board, good impact factor (number of citations)

#### Top- and mid-tier conferences

Backed by a professional organization (e.g. ACM, IEEE), peer-reviewed, archival proceedings, established, annual

#### Monographs

If published by a reputable publisher, and written by an authority on the topic (check their scholar profile!)



Edited books

Overview papers are often good; other work may be derivative (find the original paper instead)

Dissertations

From good institutions; sometimes the results and discussion are a bit "raw"

Workshop papers, posters, and mid-tier conferences Usually work-in-progress



Technical reports

- Unpublished work; usually under review... could have serious flaws!
- Non-peer reviewed papers and abstracts; presentations Not reviewed; work-in-progress, so hard to evaluate
- Work published in predatory journals/proceedings Some of them are plain gibberish



Required resources

E.g. special equipment (eye tracker, ECG, etc.)

#### Special populations

- Getting access; in person vs. online
- How long will the research take?
  - One-shot versus longitudinal study
- Ethical constraints?
  - Problematic intervention; deception



# Quality

What makes a research question "good"?



Remember: the goal of HCC is to build an organized body of knowledge about how people interact with technology

A good research question has the potential to expand this body of knowledge

- It must be grounded in the existing knowledge base
- It must be researchable
- It must be important



**Descriptive:** describe a certain phenomenon What motivates people to quit Facebook?

**Univariate:** questions pertaining to a single variable How many people have quit Facebook?

**Multivariate/correlative:** questions pertaining relationships between multiple variables

Are married people more likely to quit Facebook than single people?

How has the rate of quitting changed over time since Facebook's inception?



**Causal:** how one variable influences the other

- Does the new Facebook app increase usage among existing Facebook users compared to the old app?
- Does a comic-based privacy policy increase privacy awareness compared to a text-based policy?



Qualitative studies investigate non-numerical phenomena

- E.g., open-answer surveys, interviews, small-scale observation studies
- Usually descriptive and exploratory
- Quantitative studies investigate numerical phenomena
  - E.g., closed-answer surveys and large-scale observation studies these are univariate or multivariate
  - E.g., experiments these are causal



"Is my new travel system good?"



#### What does **good** mean?

- Learnability? (e.g. number of errors?)
- Efficiency? (e.g. time to task completion?)
- Usage satisfaction? (e.g. usability scale?)
- Outcome quality? (e.g. survey?)

We need to define **measures** 



### "Does the user interface of my travel system score high on this usability scale?"



#### What does **high** mean?

- Is 3.6 out of 5 on a 5-point scale "high"?
- What are 1 and 5?
- What is the difference between 3.6 and 3.7?
- We need to **compare** the UI against something



## "Does the UI of my system score high on this usability scale compared to this other system?"



A good research question identifies a:

- Population
- Intervention (causal) or exposure (multivariate)
- Comparative / control (causal and multivariate)
- Outcome

Note: sometimes these are implied!



Does the new Facebook app (I) increase usage (O) among existing Facebook users compared to the old app (C)?

Does a comic-based privacy policy (I) increase privacy awareness (O) compared to a text-based policy (C)? What is P here?

Are married people (I) more likely to quit Facebook (O) than single people (C)?

- l is an "exposure" here
- Implied: among Facebook users (P)



- How has the rate of quitting (O) changed over time (I) since Facebook's inception (C)?
  - Time is an "exposure" here
  - Implied: among Facebook users (P)
- What motivates people (P) to quit Facebook (O)? Descriptive, so no I and no C
- How many people (P) have quit Facebook (O)? Univariate, so no I and no C



# Brainstorning How do I find good research questions?



Write down three questions that have been on your mind related to HCC

Select one that you really would like to know the answer to

Identify why it is interesting and worth investigating

Why does it excite you? What will you gain by answering the question? What will "the field" gain (intellectual merit)? What will society gain (broader impacts)?

Is it a simple question or does it have multiple parts? Identify the parts, if any



Talk to your neighbor for 5 minutes about your question

Can you convince them that this question is interesting?

#### Together, identify the following:

- Is it about people's opinions, behaviors, or both? Why?
- Is this a qualitative or quantitative question? Why?
- Would you use experimentation or observation to answer this question? Why?
- Is this exploratory or confirmatory? Why?
- What are potential issues with answering this question?



## ACM assignment Doing a literature search



Do a preliminary search for literature on your research topic in the ACM digital library:

Conduct three (3) search queries related to your research topic

If you get too many results, try using quotation marks (e.g., "privacy paradox")



For each query:

- Write down the query
- Indicate how many search results the system returned
- Take a screenshot of the first page of results
- Read the **abstracts** of first 5 papers listed:
  - Indicate which ones you think are relevant to your topic, and why
  - Note: a paper does not have to fit your topic 100% to be relevant! Indicate which ones you think are irrelevant, and why



Finally, read the **most relevant paper** you find among the 15 papers you inspected, and explain if and how it does any of the following:

- Provide a context for your research
- Avoid duplication effort
- Argue relevance of your work
- Find relevant theories
- Identify potential problems in conducting the research
- Identify "acceptable" practices of the field