Research Design

Research Methodology in Computer Science
CSCI 514

So, You Have a Research Plan?

- You have created a new algorithm / tool / process / programming language / system / … and are looking to evaluate it.
  - How to draw conclusions? What constitutes sufficient evidence?
- You are trying to understand how software engineers / designers / … work and what challenges they face.
  - How to collect data? How to analyze it?
- You have identified some challenges (e.g., working more productively) and are looking to inform the design of possible solutions.
  - How to collect evidence that approach A is better than B?

A diversity of methods are available.

Research Exists Across Multiple Continua

Inductive ⟷ Deductive

Qualitative ⟷ Quantitative

The (Objectivist) Deductive Approach to Research

- Traditional form of research, also referred to as the scientific method, or empirical science

- Assumptions:
    - (1) there is an external reality (i.e., a real world that exists independent of the researcher)
    - (2) reality can be understood by collecting objective, unbiased data about that reality

- Knowledge builds by developing increasingly better understandings of, and insights into, the causal workings of the world

The (Objectivist) Deductive Approach to Research

- Reductionistic in nature (top-down):
  - From general, abstract conceptualizations to observable and measurable data in some context
- Research questions: testing a cause-and-effect relationship underpinning a phenomenon
- Typical approach:
  - Start with some abstract conceptualization (theory)
  - Derive a hypothesis
  - Collect data, test the hypothesis
  - Findings may falsify, support, refine, challenge, or extend the conceptualization
  - Make necessary revisions, perform additional tests
- Common method: experiments

The (Subjectivist) Inductive Approach to Research

- Assumptions:
  - (1) reality is socially and experientially constructed (i.e., reality exists because individuals and social groups share interpretations and understandings of reality);
  - (2) to understand reality, researchers need to explore the meanings constructed by individuals and groups.
- Constructivist in nature (bottom-up):
  - From specific data to a general or abstract conceptualization of the phenomenon (theory).
- Knowledge is subjective.
- Collecting data from a multitude of perspectives gives a richer and more nuanced understanding of the phenomenon.

The (Subjectivist) Inductive Approach to Research

- Research questions: explore phenomena to increase our understanding of them.
- Typical approach:
  - Start with a desire to understand or explain a particular phenomenon.
  - Collect data of and/or about this phenomenon.
  - Search for patterns across the data to generate an understanding of the phenomenon.
- Common methods: interviews, focus groups, observations.

The Types of Beliefs Held by Researchers ("Philosophical Worldviews") Influence the Practice of Research and Choice of Methods

- Positivist (or "Post-positivist") ← Objectivist Deductive
  - Knowledge is objective
  - "Causes determine effects/outcomes"
  - Reductionist: study complex things by breaking down to simpler ones
  - Prefer quantitative approaches
  - Verifying (or Falsifying) theories

- Constructivist / Interpretivist ← Subjectivist Inductive
  - Knowledge is socially constructed
  - Truth is relative to context
  - Theoretical terms are open to interpretation
  - Prefer qualitative approaches
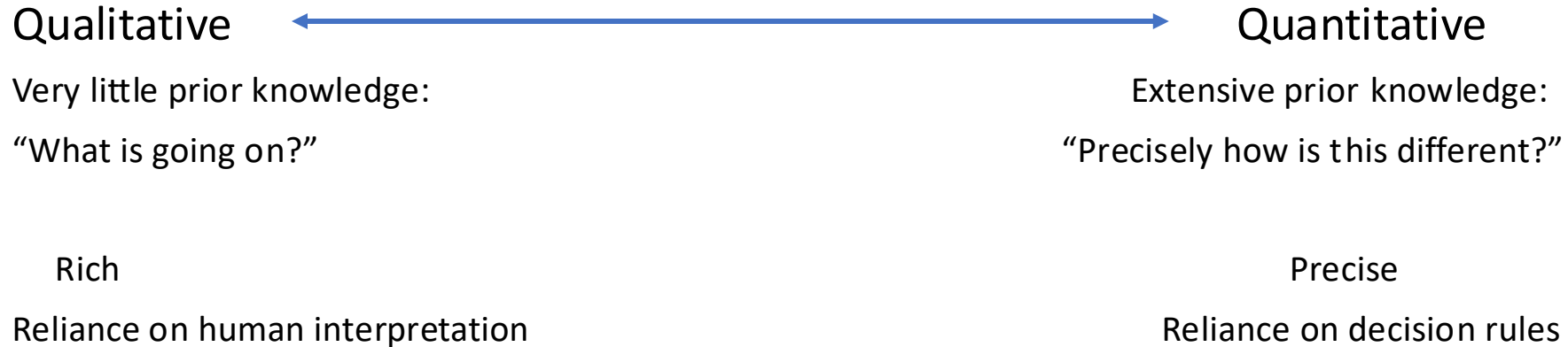 - Generating "local" theories

Four Philosophical Worldviews are Commonly Encountered

- Positivist (or "Post-positivist")
  - Knowledge is objective
  - "Causes determine effects/outcomes"
  - Reductionist: study complex things by breaking down to simpler ones
  - Prefer quantitative approaches
  - **Verifying (or Falsifying) theories**

- Constructivist / Interpretivist
  - Knowledge is socially constructed
  - Truth is relative to context
  - Theoretical terms are open to interpretation
  - Prefer qualitative approaches
  - **Generating "local" theories**

- Transformative/Advocate/Critical Theorist
  - Research is political act
  - Knowledge is created to empower groups/individuals
  - Prefer participatory approaches
  - **Seeking change in society**

- Pragmatist
  - Research is problem-centered
  - Prefer multiple methods / multiple perspectives
  - **Seeking practical solutions to problems**

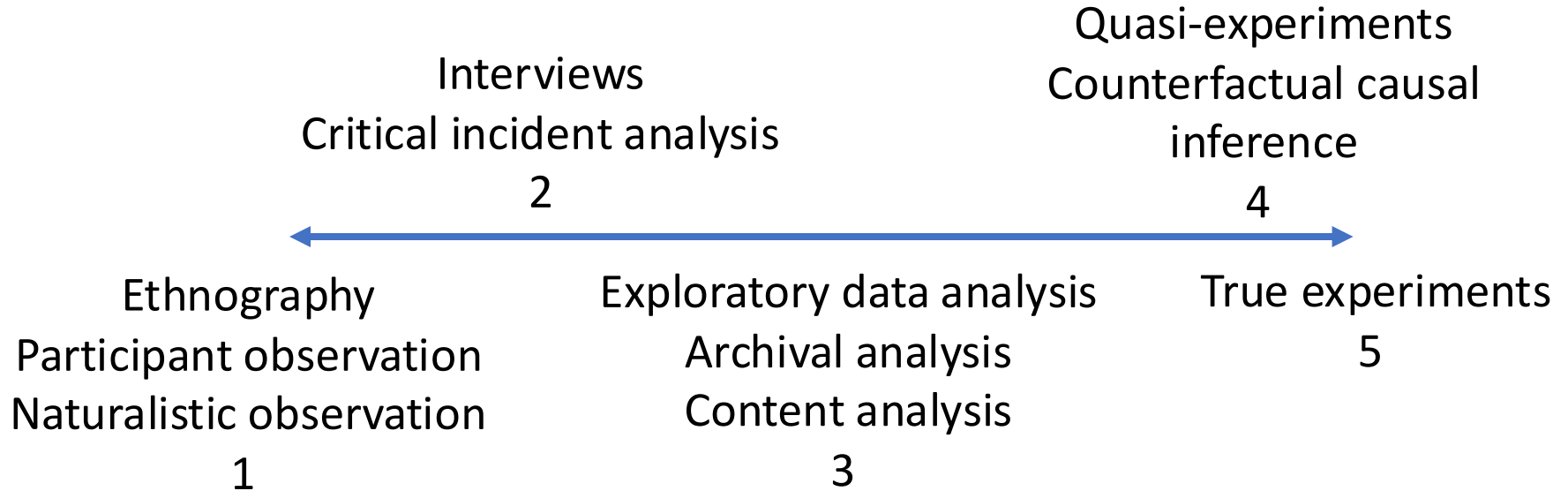# Selection of Methods Depend on

- Approach to Research
- Nature of Contribution
- Specific Research Question
- State of Knowledge

Another Few Ways of Thinking About How Methods Are Related

Qualitative <————————————————————> Quantitative

Very little prior knowledge: Extensive prior knowledge:

"What is going on?"   "Precisely how is this different?"

Rich   Precise

Reliance on human interpretation   Reliance on decision rules

Often most effective to use methods in combination or in sequence (rich to precise)

Points on the Spectrum

Interviews
Critical incident analysis
2

Quasi-experiments
Counterfactual causal inference
4

Ethnography
Participant observation
Naturalistic observation
1

Exploratory data analysis
Archival analysis
Content analysis
3

True experiments
5

Activity 1

- Select 5 papers from your ongoing literature review
- For each paper, complete the following table:

Paper Citation	Research Method	Quantitative / Qualitative / Mixed?	Type of Data Collected	Data Analysis Approach
e.g., Smith et al. (2021)	Controlled Experiment	Quantitative	Task completion times, error rates	Statistical analysis (ANOVA)

After completing the table, write a short (200–300 word) reflection:

- What methods are most common in your set?

Disclaimer: All Methods are Flawed

- Laboratory experiments
  - Cannot study large scale software development in the lab!
  - Too many variables to control them all!
- Case studies
  - How do we know what's true in one project generalizes to others?
  - Researcher chose what questions to ask, hence biased the study
- Surveys
  - Self-selection of respondents biases the study
  - Respondents tell you what they think they ought to do, not what they actually do
- etc…

Each method has its own standards and techniques for rigor.
When applied correctly, they can be useful.

Activity 2

Review the following slides to learn about some of the research methods applications and their pros and cons, then answer:

- Which method(s) might be appropriate for your own research question? Why?

- What research questions could your chosen method(s) help you answer in your study?

- What limitations of the method(s) concern you? How might you mitigate them in your design?

- What resources or constraints do you anticipate in using this method? (e.g., access to participants, time, tools)

An Ethnographic Study of Copy and Paste Programming Practices in OOPL

Miryung Kim[1]   Lawrence Bergman[2]   Tessa Lau[2]   David Notkin[1]

Department of Computer Science & Engineering
University of Washington[1]
{miryung, notkin}@cs.washington.edu

IBM T. J. Watson Research Center[2]
{bergmanl, tessalau}@us.ibm.com

Abstract

Although programmers frequently copy and paste code when they develop software, implications of common copy and paste (C&P) usage patterns have not been studied previously. We have conducted an ethnographic study in order to understand programmers' C&P programming practices and discover opportunities to assist common C&P usage patterns. We observed programmers using an instrumented Eclipse IDE and then analyzed why and how they use C&P operations. Based on our analysis, we constructed a taxonomy of C&P usage patterns.

This paper presents our taxonomy of C&P usage patterns and discusses our insights with examples

locate code duplicates and refactor existing duplications to a unit of programming language abstraction [1][2][3][6][12][14][15]. However, in practice, a substantial amount of duplicated code is still present in many software systems [6][12]. Our understanding of how and why code clones are created is very limited.

Earlier studies have formed a few informal hypotheses about how C&P is performed by programmers to reuse code [17][18]. However, existing work has not focused specifically on solving the possible problems that can be incurred by C&P during software evolution.

The main purpose of our work is to investigate common C&P usage patterns and associated

Example: Ethnography Rich Precise

- What is it?
  - Immersion in the environment, group
  - Attempt to see the world through their eyes

- What questions can it answer?
  - How do the participants think about their work?
  - What are the problems?

- What makes it rigorous?
  - Constantly testing interpretations
  - Triangulation – multiple sources of data

Example: Ethnography Rich Precise

- What contributions can it support?
  - Problem as seen by persons of interest
  - Work in context
- What are its limitations?
  - May get trapped by participants' perceptions
  - Small samples, no causality
- What resources are needed?
  - Time and labor intensive
  - Access to right people
  - Willingness and ability to join group

Understanding Mental Ill-health as Psychosocial Disability: Implications for Assistive Technology

Kathryn E. Ringland[1,A,B], Jennifer Nicholas[1,B], Rachel Kornfield[A,B], Emily G. Lattie[B], David C. Mohr[B], Madhu Reddy[A]

[A] Department of Communication Studies
Northwestern University
Chicago, Illinois, USA
{kathrynringland, rachel.kornfield, mreddy}@northwestern.edu

[B] Center for Behavioral Intervention Technologies
Department of Preventive Medicine
Northwestern University
Chicago, Illinois, USA
{j.nicholas, emily.lattie, d-mohr}@northwestern.edu

ABSTRACT

Psychosocial disability involves actual or perceived impairment due to a diversity of mental, emotional, or cognitive experiences. While assistive technology for psychosocial disabilities has been understudied in communities such as ASSETS, advances in computing have opened up a number of new avenues for assisting those with psychosocial disabilities beyond the clinic. However, these tools continue to emerge primarily within the framework of "treatment," emphasizing resolution or improvement of mental health symptoms. This work considers what it means to adopt a social model lens from disability studies and incorporate the expertise of assistive technology researchers in relation to mental health. Our investigation draws on interviews conducted with 18 individuals who have complex health needs that include mental health symptoms. This work highlights the potential role for assistive technology in supporting psychosocial disability outside of a clinical or medical framework.

diversity of mental, emotional, or cognitive experiences. Individuals with these disabilities also represent a large portion of the population, with 1 in 4 people diagnosed with a mental health condition worldwide [56]. In addition, many individuals may not meet the medical "threshold" for mental health conditions but may still find their daily living activities impaired by mental health symptoms [53]. Psychosocial disabilities also frequently go hand-in-hand with physical health conditions and disabilities [70], and this co-occurrence can produce a more complex experience of disability relative to those who have only mental or physical health concerns [90]. However, holistic support that addresses both physical and psychosocial disabilities is widely underdeveloped and understudied.

While individuals with disability still broadly face issues of stigma and oppression, psychosocial disabilities in particular are complicated by their invisibility and unique place in history [36]. Medical treatment of mental health conditions has historically neglected consideration of patients'

Example: Interviews Rich Precise

- What is it?
  - Structured interaction
  - Questions, answers, followup
- What questions can it answer?
  - Perceptions, opinions, processed observations
  - How things are done, exceptions, problems
- What makes it rigorous?
  - Preparation with well thought-out topics
  - Cross-validation in questions, interviewees, checking interpretations

Example: Interviews Rich Precise

- What contributions can it support?
  - Nature of problem, as perceived, current process
  - Examples, exceptions, incidents
- What are its limitations?
  - Information is processed, filtered by interviewees
  - May be inappropriately biased by questions
- What resources are needed?
  - Willing interviewees, correctly positioned
  - Ability to sample all relevant perspectives
  - Preparation, follow-up

How Work From Home Affects Collaboration: A Large-Scale Study of Information Workers in a Natural Experiment During COVID-19

Longqi Yang, Sonia Jaffe, David Holtz, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, CJ Lee, Brent Hecht, Jaime Teevan

Microsoft Corporation
Correspondence: Longqi.Yang@microsoft.com

ABSTRACT

The COVID-19 pandemic has had a wide-ranging impact on information workers such as higher stress levels, increased workloads, new workstreams, and more caregiving responsibilities during lockdown. COVID-19 also caused the overwhelming majority of information workers to rapidly shift to working from home (WFH). The central question this work addresses is: can we isolate the effects of WFH on information workers' collaboration activities from all other factors, especially the other effects of COVID-19? This is important because in the future, WFH will likely to be more common than it was prior to the pandemic.

We use difference-in-differences (DiD), a causal identification strategy commonly used in the social sciences, to control for unobserved confounding factors and estimate the causal effect of WFH. Our analysis relies on measuring the difference in changes between those who WFH prior to COVID-19 and those who did not. Our preliminary results suggest that on average, people spent more time on collaboration in April (Post WFH mandate) than in February (Pre WFH mandate), but this is primarily due to factors other than WFH, such as lockdowns during the pandemic. The change attributable to WFH specifically is in the opposite direction: less time on collaboration and more focus time. This reversal shows the importance of using causal inference: a simple analysis would have resulted in the wrong conclusion. We further find that the effect of WFH is moderated by individual remote collaboration experience prior to WFH. Meanwhile, the medium for collaboration has also shifted due to WFH: instant messages were used more, whereas scheduled meetings were used less. We discuss design implications -- how future WFH may affect focused work, collaborative work, and creative work.

KEYWORDS

Work-from-home, remote work, COVID-19, causal inference, information worker

---

Yang, Longqi, et al. "How work from home affects collaboration: A large-scale study of information workers in a natural experiment during COVID-19." *arXiv preprint arXiv:2007.15584* (2020).

Example: Quasi-Experiment Rich Precise

- What is it?
  - Naturally-occurring differences
  - Examination of effects of variables in situ

- What questions can it answer?
  - What are the effects of introducing X?
  - What is the difference between X and Y?

- What makes it rigorous?
  - Good quasi-control groups
  - Access to data for control variables

Example: Quasi-Experiment Rich Precise

- What contributions can it support?
  - Value of tool, method, process, training
  - Influence of context factors

- What are its limitations?
  - Never sure cause-effect relation is established
  - Relying on luck, that situation occurs naturally

- What resources does it take to do it well?
  - Sophisticated statistics (e.g., multiple regression)
  - Contextual knowledge of experimental situation

Development and Evaluation of a Tool for Assisting Content Creators in Making PDF Files More Accessible

DEBASHISH PRADHAN, TRIPTI RAJPUT, ARAVIND JEMBU RAJKUMAR, and
JONATHAN LAZAR, College of Information Studies, Trace Research and Development Center,
and Human-Computer Interaction Lab (HCIL), University of Maryland
RAJIV JAIN, VLAD I. MORARIU, and VARUN MANJUNATHA, Adobe Research

3

Most PDF documents are inaccessible for people with disabilities, creating barriers in education, science, commerce, e-government, and recreation. Documents in PDF format are considered harder to make accessible than documents in other formats, primarily due to the insufficient tools available to assist content creators. In this article, we present the research and development of Ally, a new tool to assist content creators in remediating their PDF files to improve accessibility. Ally utilizes best practices from other areas of HCI research to create a more efficient and effective interaction for remediating regions, headers, reading order, and tables in a PDF document for improved accessibility. Twenty participants attempted to complete the same PDF accessibility remediation tasks using both Ally and a standard industry tool, Adobe Acrobat Pro. Ally was almost twice as fast and three times as accurate compared to Acrobat Pro, with participants reporting a strong preference for and a much higher level of satisfaction with Ally. The approaches taken in Ally improve the ability to create accessible PDFs efficiently and accurately for the four important aspects studied, but future work will need to incorporate additional functionality, related to remediating alt text, forms, and other aspects of PDF accessibility.

CCS Concepts: • **Human-centered computing** → **Accessibility systems and tools**; **Interactive systems and tools**;

Additional Key Words and Phrases: Portable Document Format (PDF), PDF accessibility, content accessibility, PDF universal access, matterhorn protocol

Pradhan, Debashish, et al. "Development and Evaluation of a Tool for Assisting Content Creators in Making PDF Files More Accessible." *ACM Transactions on Accessible Computing (TACCESS)* 15.1 (2022): 1-52.

Example: True Experiment Rich Precise

- What is it?
  - Comparison that is engineered
  - Random assignment of values of independent variables

- What questions can it answer?
  - Cause and effect
  - Size of effect, interaction of factors

- What makes it rigorous?
  - Well designed experimental and control conditions
  - Attention to reliability, validity

Example: True Experiment Rich Precise

- What contributions can it support?
  - Value of tool, method, process, training
  - Influence of context factors

- What are its limitations?
  - Must be able to isolate critical variables
  - Limited by ability to create situations that manipulate people

- What resources does it take to do it well?
  - Access to appropriate subjects
  - Statistics, measurement instruments