Tools for thinking about research

Dr. Chat Wacharamanotham
University of Zurich

29 May 2025

Use your laptop

https://chatw.ch/research-thinking

Chatchavan Wacharamanotham

ชัชวาล วัชรมโนธรรม

Lecturer at the Department of Informatics, University of Zurich, Switzerland

Previously: Assistant professor at University of Zurich; Lecturer at Swansea University (UK); PhD in Human–Computer interaction from RWTH Aachen University

Research: Improving how computers can help people do better, and transparent science

Past research: Interaction techniques for touch input on and above screens

Use your laptop

https://chatw.ch/research-thinking

Chatchavan Wacharamanotham

ชัชวาล วัชรมโนธรรม

Works in the field of HCI since 2010

- 7 papers at the CHI (3 best paper awards; top 1%)
- 16 years reviewing for CHI (13 special recognitions on the reviews)

Roles in the CHI conferences: Associate Chair (2022–23) • Best Paper Award Committee (2022) • Student Research Competition Co-chair (2023)

Roles in journals: Associate editor of IJHCS (International Journal of Human-Computer Studies) • Organizer of JoVI (The Journal of Visualization and Interaction)

Use your laptop

https://chatw.ch/research-thinking

Chatchavan Wacharamanotham

ชัชวาล วัชรมโนธรรม

Roles in the HCI communities in Thailand and Asia:

- Co-founded of the Bangkok SIGCHI Chapter (2017)
- Co-organized Asian CHI Symposia (2019–2020)
- Chaired the ACM SIGCHI Asia Committee (2024)

Thoughts:

- Evaluation methods for HCI-related research in Thailand
- Justifying conference publication as a valid outlet for HCI research
- Promoting HCI research: method, funding, and communities in Thailand

Use your laptop

https://chatw.ch/research-thinking

Goals

Questions that you should be able to answer at the end of this talk

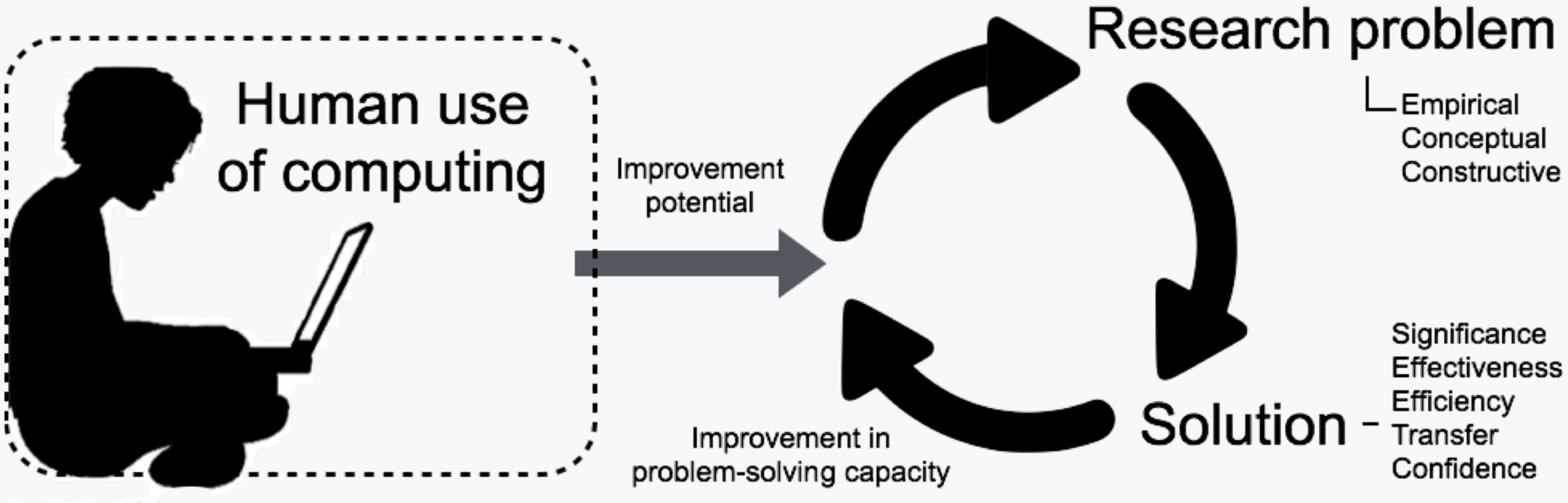
- What conceptual tools that can help me understand research?
- What are ways to validate design/engineering research (beyond questionnaires)?

Use your laptop

https://chatw.ch/research-thinking

Research as problem-solving

Research as problem-solving



Oulasvirta, A., & Hornbæk, K. (2016, May). HCI research as problem-solving. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4956-4967).

Research **problems** in HCI

Empirical 🔍
- Unknown phenomena
- Unknown factors
- Unknown effects

Constructive 🔧
- no known solution
- partial, ineffective, or inefficient solution
- insufficient knowledge or resources for implementation or deployment

Conceptual 💭
- Implausibility
- Inconsistency
- Incompatibility

Example research problems about "AI fairness"

🔍 Which AI application are unfair?

🔍 Who are affected?

🔍 What causes it?

🔧 What are possible ways to solve AI unfairness?

🔧 Which ways are cost-effective?

🔍 💭 🔧 How to measure the effectiveness of the solutions?

Oulasvirta, A., & Hornbæk, K. (2016, May). HCI research as problem-solving. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4956-4967).

Exercise: Identify types of your research problem **Miro**

You have 10 minutes to:

1. Locate your seat on the Miro board
2. Write 1–2 sentences about your research problem
3. Add one or more tags: #empirical, #constructive, #conceptual
4. Think:
    - Why did you choose these tags?
    - Could there be other possible tags that are applicable?

~~Together with a person next to you, take 10 minutes per person:~~

- ~~Give an elevator-pitch of your research~~
- ~~Discuss your tags:~~
    - ~~Do other share your classification?~~
    - ~~What are other tags that may be applicable?~~

Types of research problem

Empirical 🔍

- Unknown phenomena
- Unknown factors
- Unknown effects

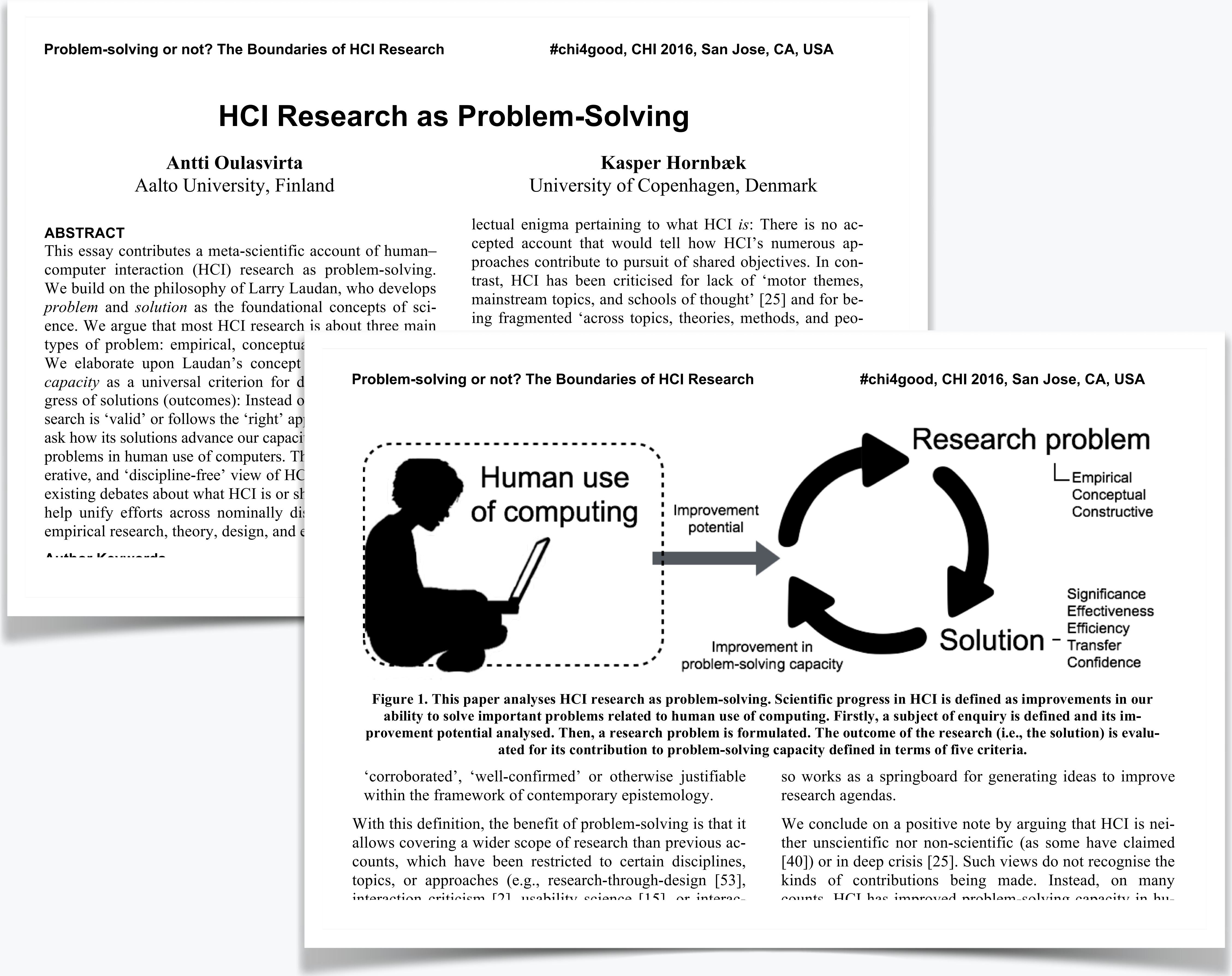
Constructive 🔧

- no known solution
- partial, ineffective, or inefficient solution
- insufficient knowledge or resources for implementation or deployment

Conceptual 💭

- Implausibility
- Inconsistency
- Incompatibility

To learn more about this framework…



Problem-solving or not? The Boundaries of HCI Research #chi4good, CHI 2016, San Jose, CA, USA

HCI Research as Problem-Solving

Antti Oulasvirta
Aalto University, Finland

Kasper Hornbæk
University of Copenhagen, Denmark

ABSTRACT
This essay contributes a meta-scientific account of human–computer interaction (HCI) research as problem-solving. We build on the philosophy of Larry Laudan, who develops *problem* and *solution* as the foundational concepts of science. We argue that most HCI research is about three main types of problem: empirical, conceptu…

…lectual enigma pertaining to what HCI *is*: There is no accepted account that would tell how HCI's numerous approaches contribute to pursuit of shared objectives. In contrast, HCI has been criticised for lack of 'motor themes, mainstream topics, and schools of thought' [25] and for being fragmented 'across topics, theories, methods, and peo-

Problem-solving or not? The Boundaries of HCI Research #chi4good, CHI 2016, San Jose, CA, USA

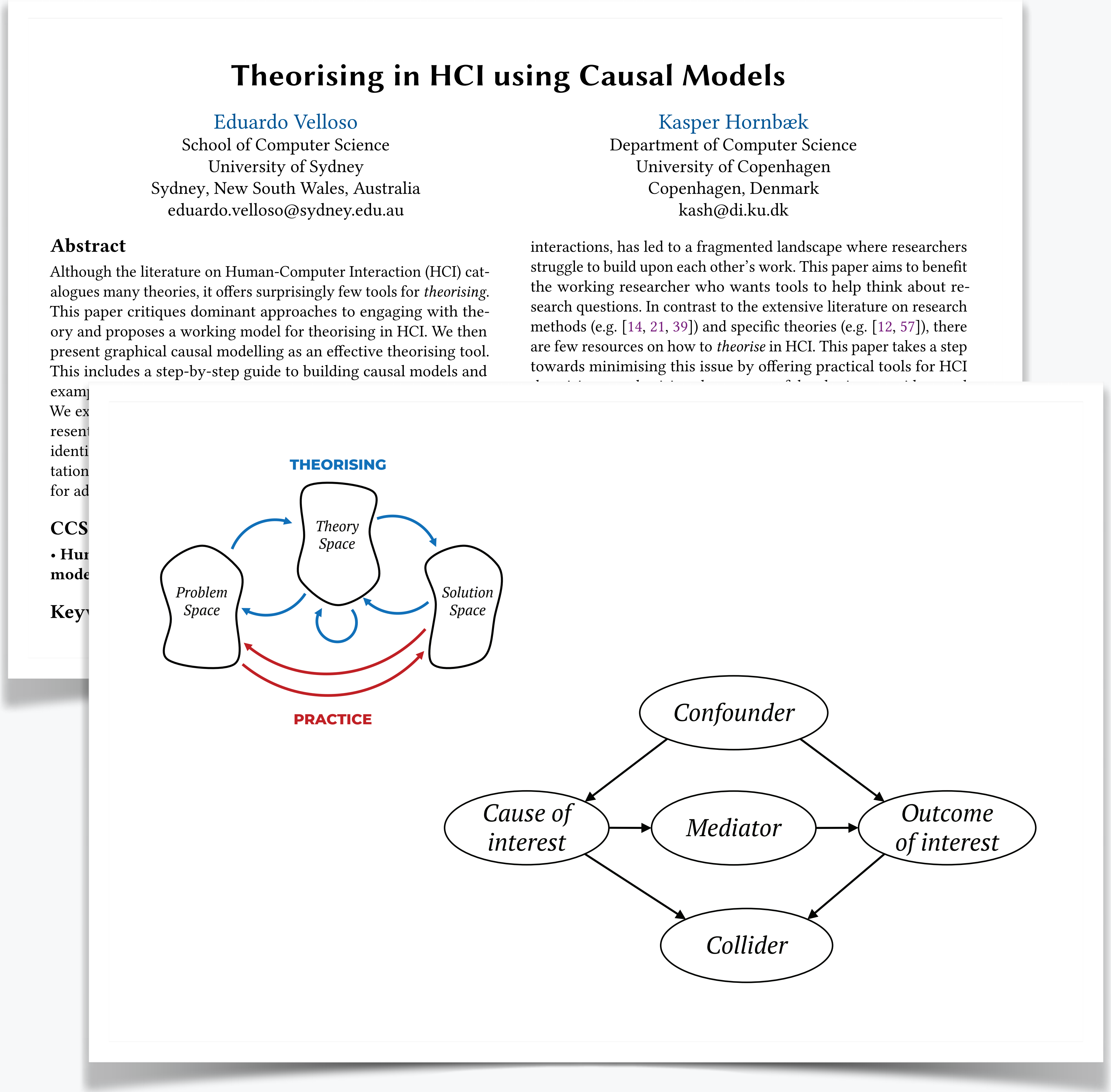
Figure 1. This paper analyses HCI research as problem-solving. Scientific progress in HCI is defined as improvements in our ability to solve important problems related to human use of computing. Firstly, a subject of enquiry is defined and its improvement potential analysed. Then, a research problem is formulated. The outcome of the research (i.e., the solution) is evaluated for its contribution to problem-solving capacity defined in terms of five criteria.

'corroborated', 'well-confirmed' or otherwise justifiable within the framework of contemporary epistemology.

With this definition, the benefit of problem-solving is that it allows covering a wider scope of research than previous accounts, which have been restricted to certain disciplines, topics, or approaches (e.g., research-through-design [53],

so works as a springboard for generating ideas to improve research agendas.

We conclude on a positive note by arguing that HCI is neither unscientific nor non-scientific (as some have claimed [40]) or in deep crisis [25]. Such views do not recognise the kinds of contributions being made. Instead, on many



Theorising in HCI using Causal Models

Eduardo Velloso
School of Computer Science
University of Sydney
Sydney, New South Wales, Australia
eduardo.velloso@sydney.edu.au

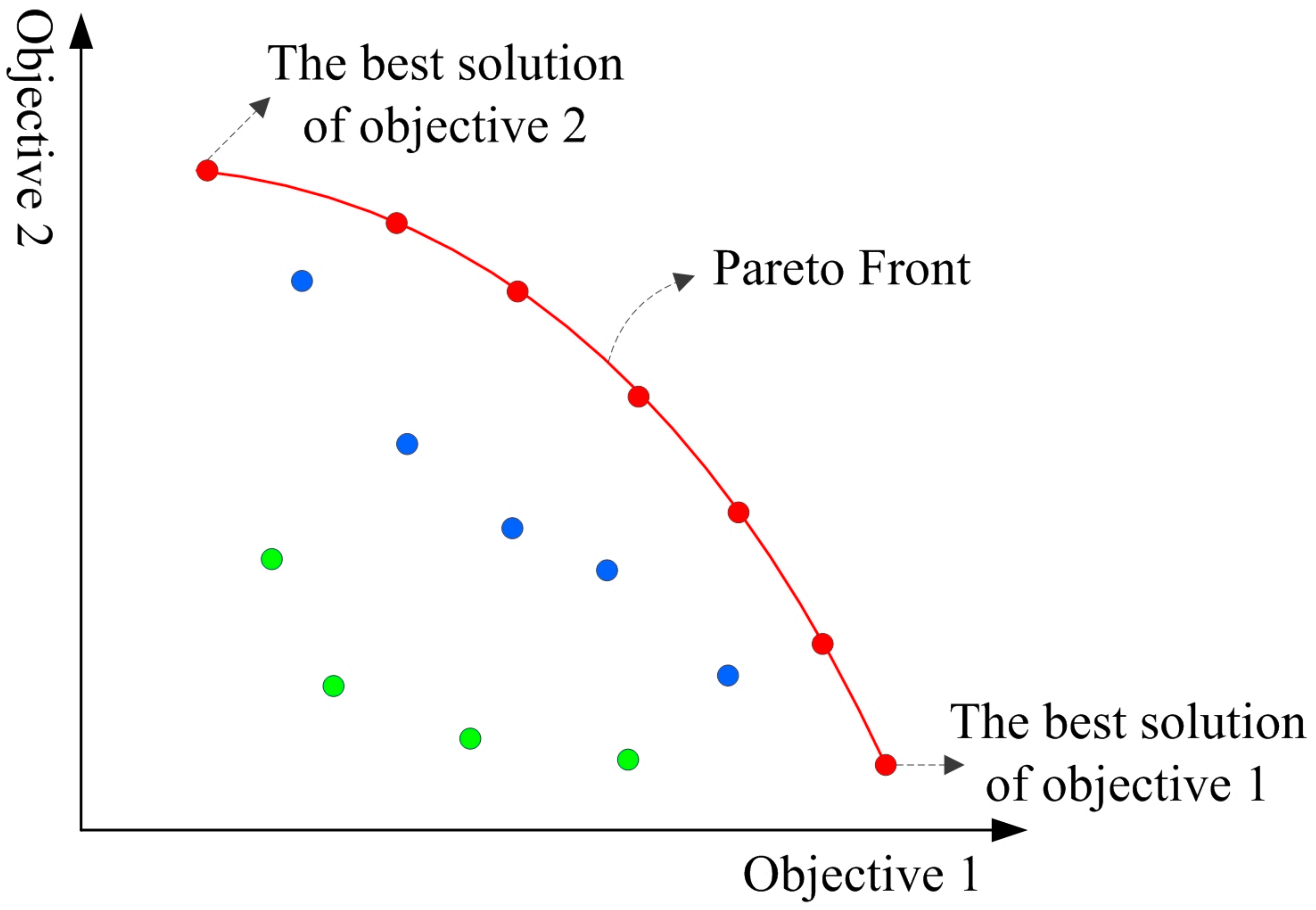
Kasper Hornbæk
Department of Computer Science
University of Copenhagen
Copenhagen, Denmark
kash@di.ku.dk

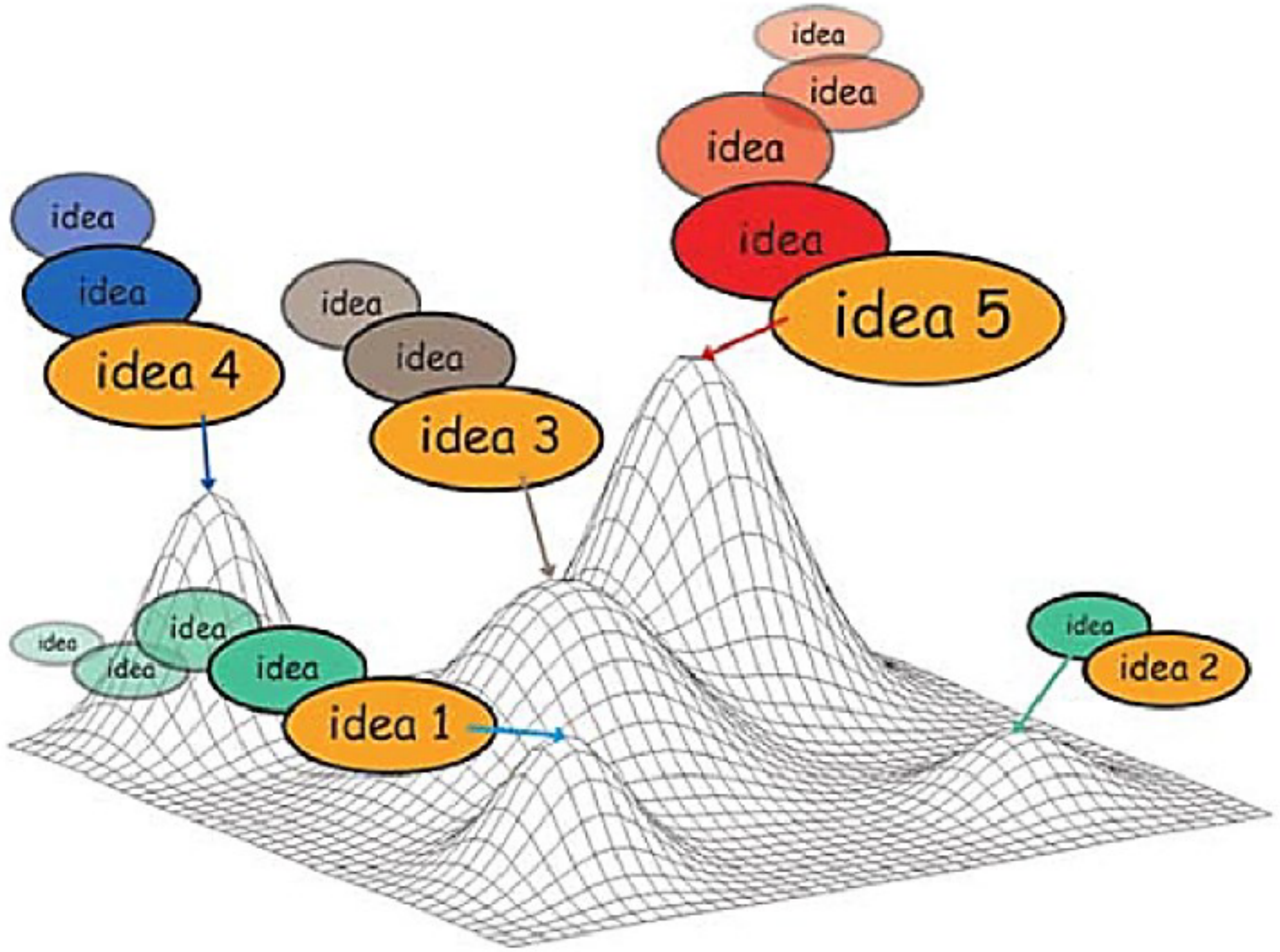
Abstract
Although the literature on Human-Computer Interaction (HCI) catalogues many theories, it offers surprisingly few tools for *theorising*. This paper critiques dominant approaches to engaging with theory and proposes a working model for theorising in HCI. We then present graphical causal modelling as an effective theorising tool. This includes a step-by-step guide to building causal models and …

interactions, has led to a fragmented landscape where researchers struggle to build upon each other's work. This paper aims to benefit the working researcher who wants tools to help think about research questions. In contrast to the extensive literature on research methods (e.g. [14, 21, 39]) and specific theories (e.g. [12, 57]), there are few resources on how to *theorise* in HCI. This paper takes a step towards minimising this issue by offering practical tools for HCI

Oulasvirta, A., & Hornbæk, K. (2016, May). HCI research as problem-solving. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4956-4967).
Velloso, E., & Hornbæk, K. (2025). Theorising in HCI using Causal Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*

Constructive research



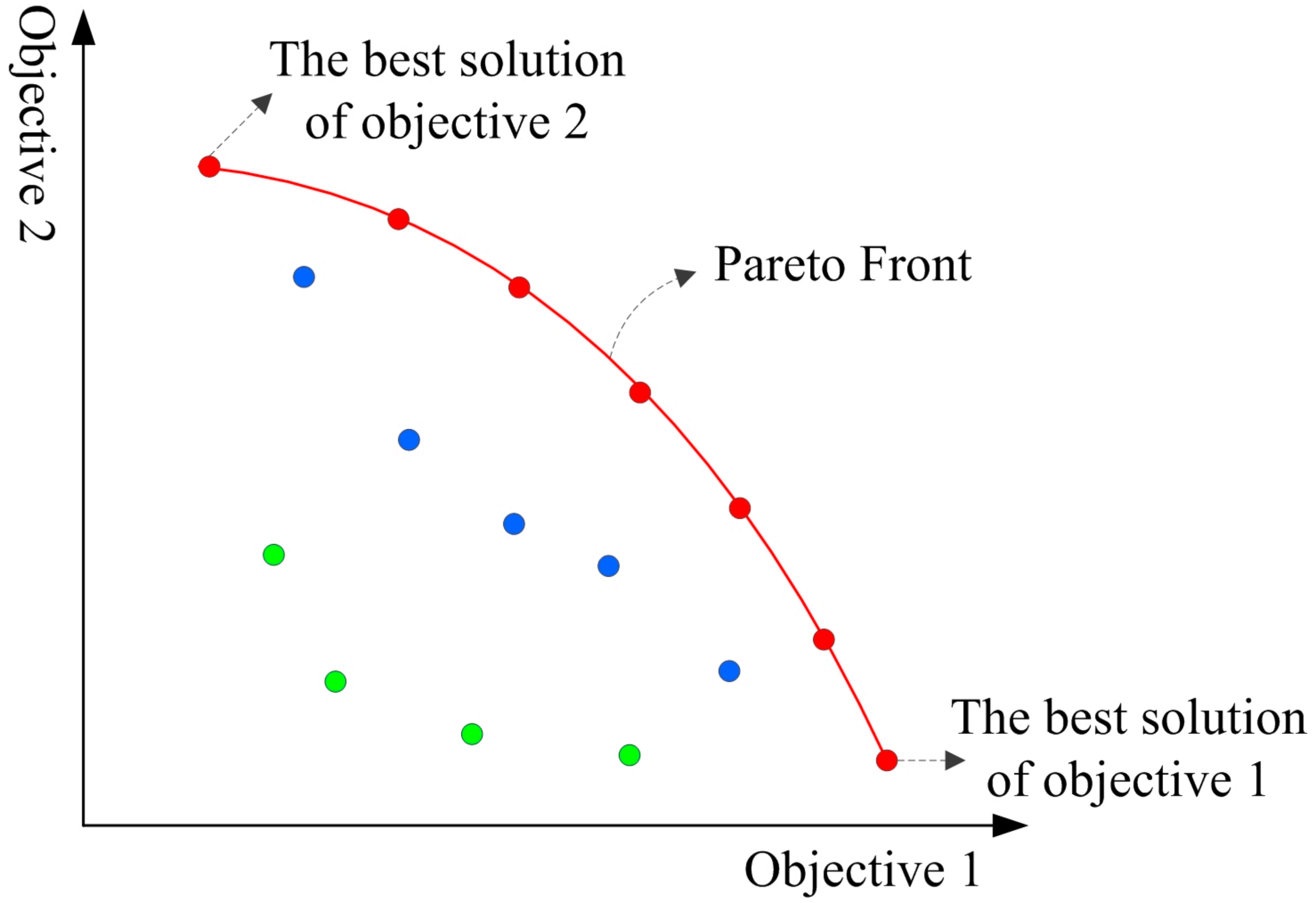


Left diagram from Yang L et al. (2018). A Knowledge-Informed and Pareto-Based Artificial Bee Colony Optimization Algorithm for Multi-Objective Land-Use Allocation. ISPRS International Journal of Geo-Information. 2018; 7(2):63.
Right diagram from Greenberg et al. (2012) Sketching User Experience — The Workbook

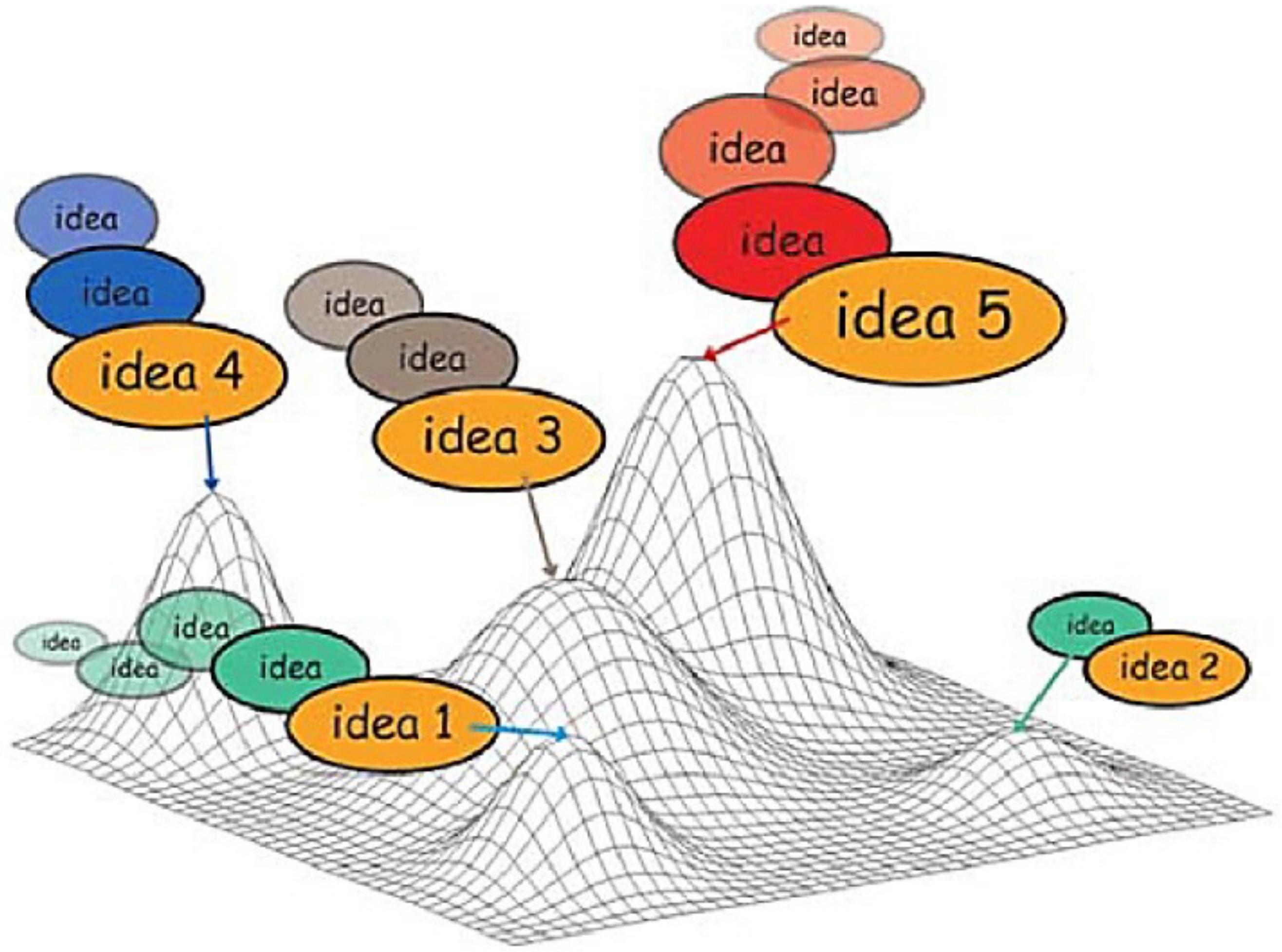
Satisficing

Producing any one of what might well be a large range of satisfactory solutions rather than attempting to generate the one hypothetically-optimum solution.

Etymology: Satisfy + sacrifice



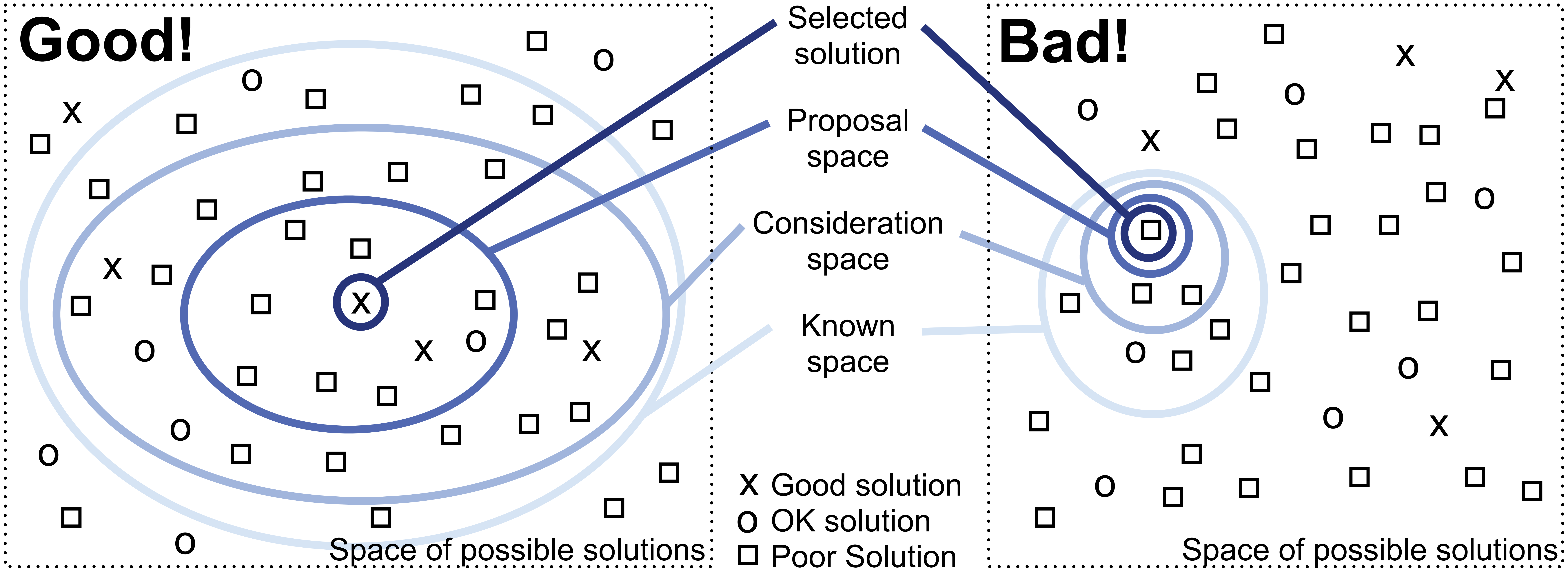
Optimizing



Satisficing

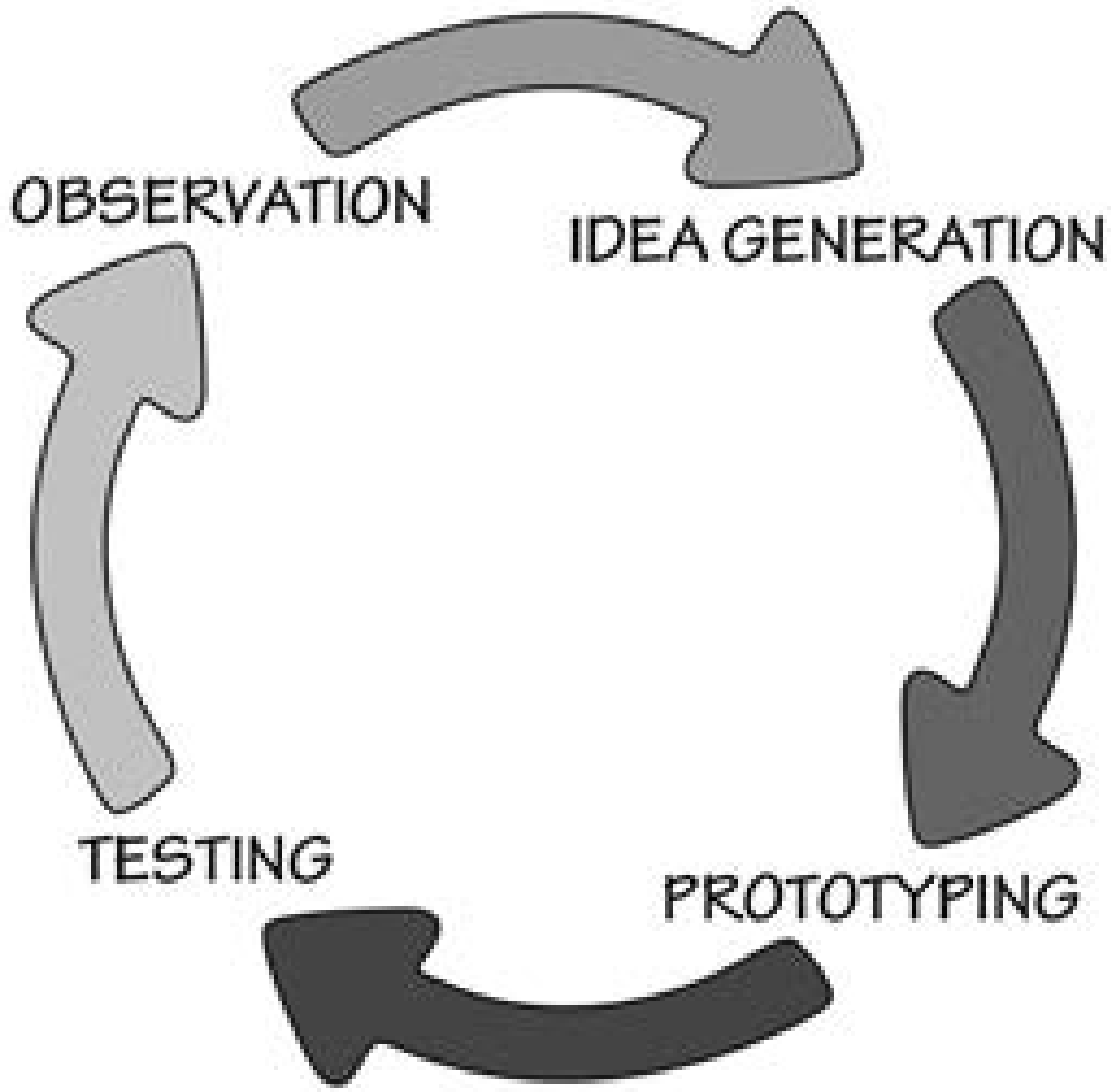
Satisficing concept: Simon, H A The sciences of the artificial MIT Press, Cambridge, MA, USA (1969)
Left diagram from Yang L et al. (2018). A Knowledge-Informed and Pareto-Based Artificial Bee Colony Optimization Algorithm for Multi-Objective Land-Use Allocation. ISPRS International Journal of Geo-Information. 2018; 7(2):63.
Right diagram from Greenberg et al. (2012) Sketching User Experience — The Workbook

Navigating the solution space



 Diagrams: Munzner (2015) Visualization analysis and design. CRC Press

Design process is iterative



- Do you like the system?
- System Usability Scale (SUS)

 Don Norman (2013) The Design of Everyday Things. Basic Books.

Q: Do you like this system?

Problems with this type of test:

Hawthorne effect: Participants behave differently because they are aware that their behaviors are measured

Social-desirability bias: People tend answer questions in a manner that will be viewed favorably by others (= you the researchers!)

	Strongly disagree				Strongly agree	
	1	2	3	4	5	
1. I think that I would like to use this system frequently					√	4
2. I found the system unnecessarily complex				√		1
3. I thought the system was easy to use		√				1
4. I think that I would need the support of a technical person to be able to use this system	√					4
5. I found the various functions in this system were well integrated		√				1

Usability study is not publishable

CHI 2008 Proceedings · Usability Evaluation Considered Harmful? April 5-10, 2008 · Florence, Italy

Usability Evaluation Considered Harmful (Some of the Time)

Saul Greenberg
Department of Computer Science
University of Calgary
Calgary, Alberta, T2N 1N4, Canada
saul.greenberg@ucalgary.ca

Bill Buxton
Principle Researcher
Microsoft Research
Redmond, WA, USA
bibuxton@microsoft.com

ABSTRACT
Current practice in Human Computer Interaction as encouraged by educational institutes, academic review processes, and institutions with usability groups advocate usability evaluation as a critical part of every design process. This is for good reason: usability evaluation has a significant role to play when conditions warrant it. Yet evaluation can be ineffective and even harmful if naively done 'by rule' rather than 'by thought'. If done during early stage design, it can mute creative ideas that do not conform to current interface norms. If done to test radical innovations, the many interface issues that would likely arise from an immature technology can quash what could have been an inspired vision. If done to validate an

INTRODUCTION
Usability evaluation is one of the major cornerstones of user interface design. This is for good reason. As Dix et al., remind us, such evaluation helps us "assess our designs and test our systems to ensure that they actually behave as we expect and meet the requirements of the user" [7]. This is typically done by using an evaluation method to measure or predict how effective, efficient and/or satisfied people would be when using the interface to perform one or more tasks. As commonly practiced, these usability evaluation methods range from laboratory-based user observations, controlled user studies, and/or inspection techniques [7,22,1]. The scope of this paper concerns these methods.

Usable ≠ useful

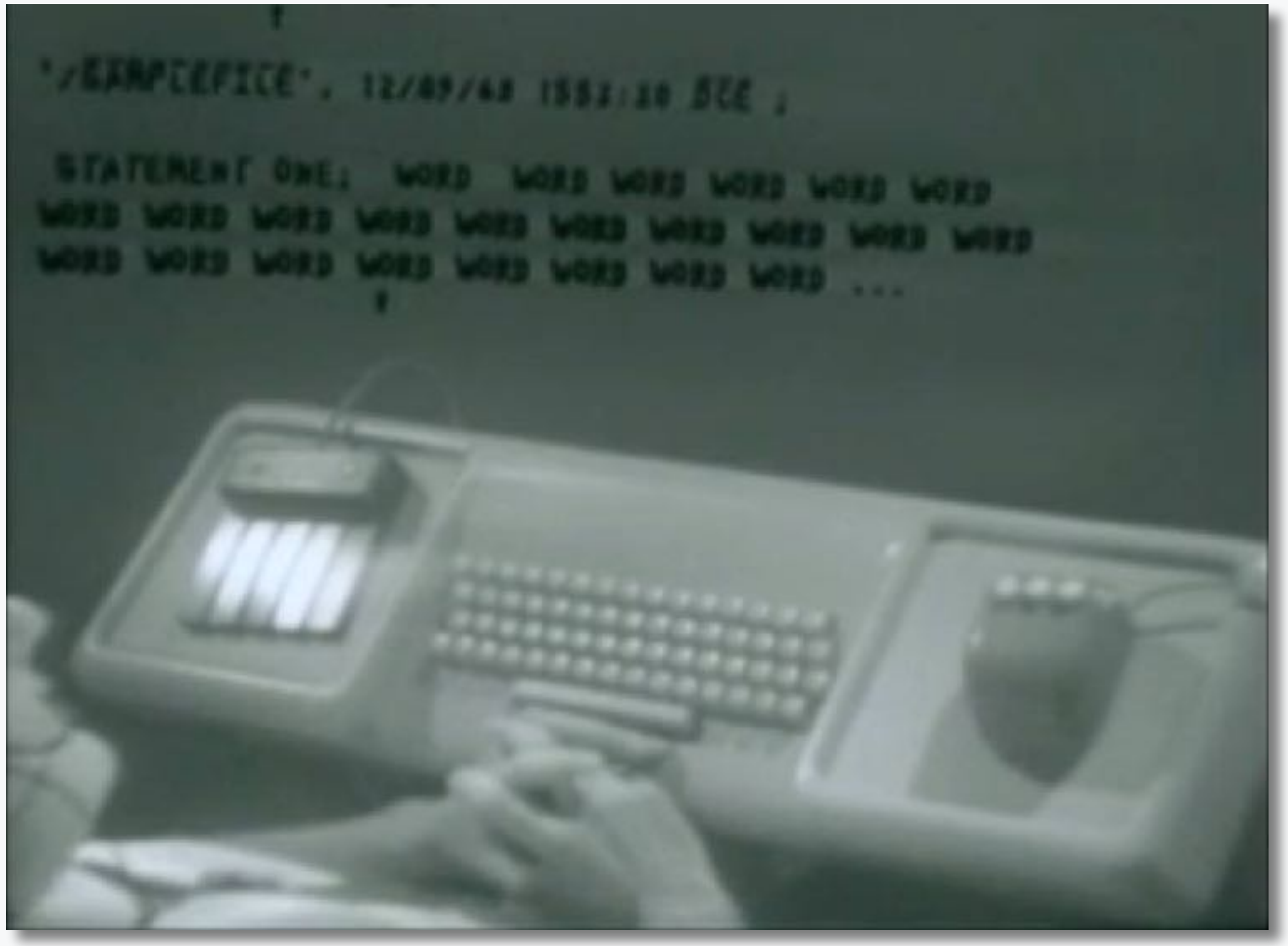
Visionary ideas shown through prototypes could impact the way people think about the problem or the solution

But the prototypes themselves may not be practical to deploy

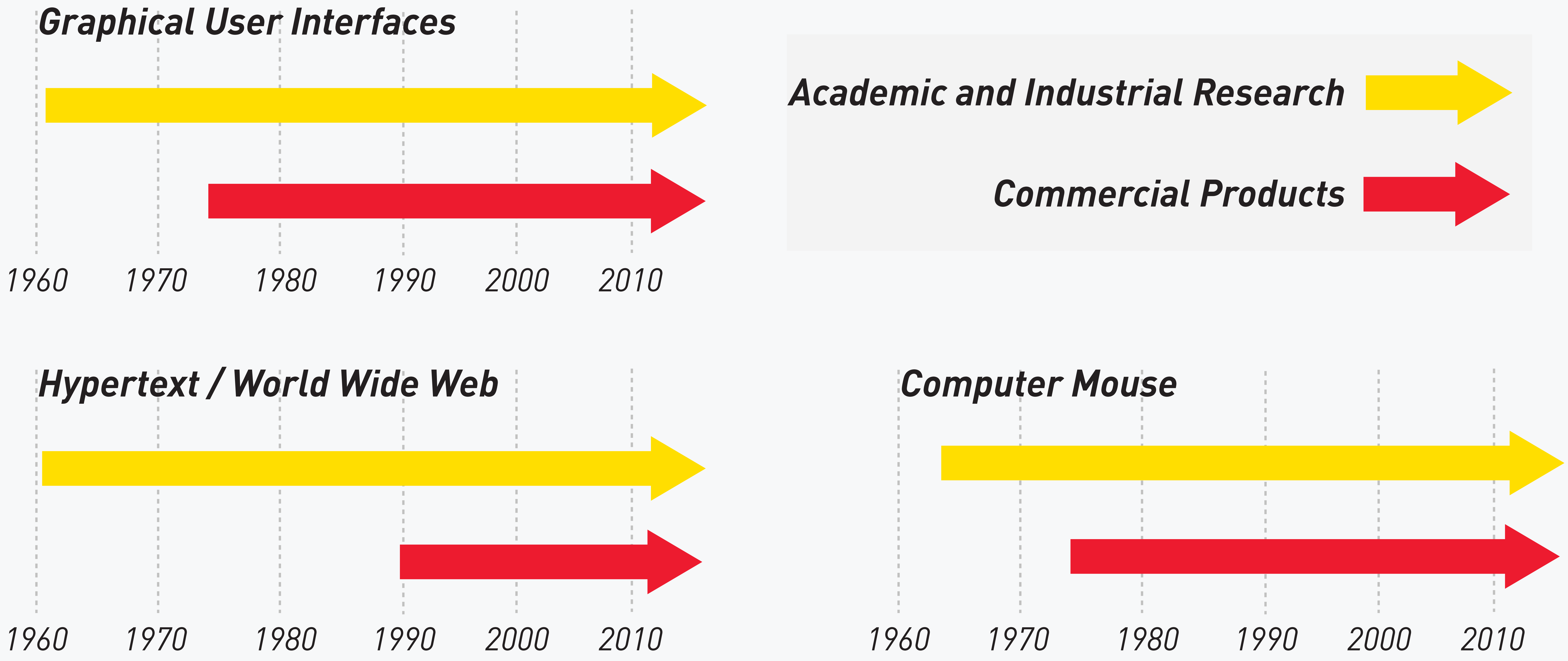
Greenberg, S., & Buxton, B. (2008, April). Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 111-120).

Mother of all demos (Douglas Engelbart 1968)

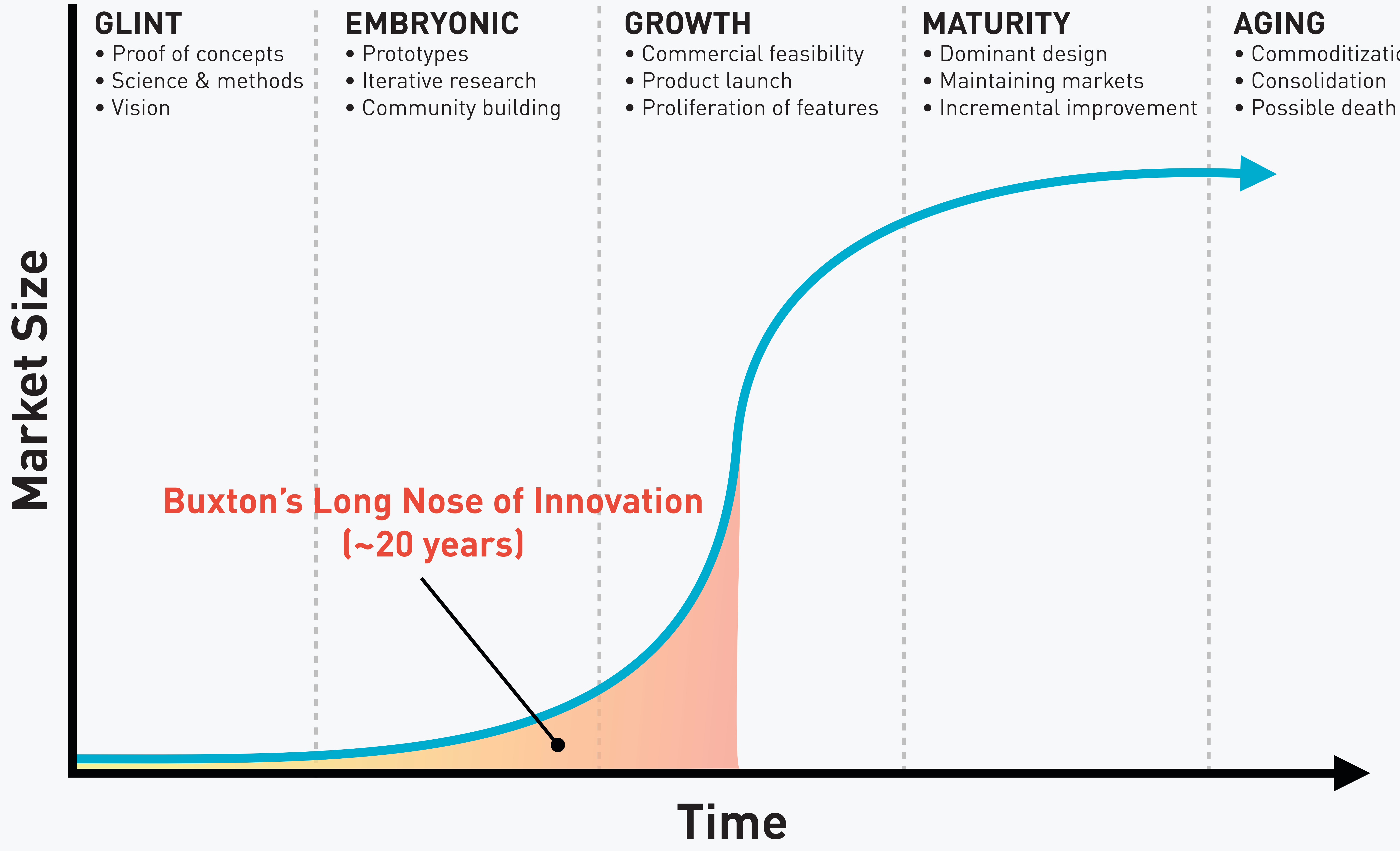
Demonstrate the NLS system with mouse input, hyper text, video conferencing, and collaborative real-time editor



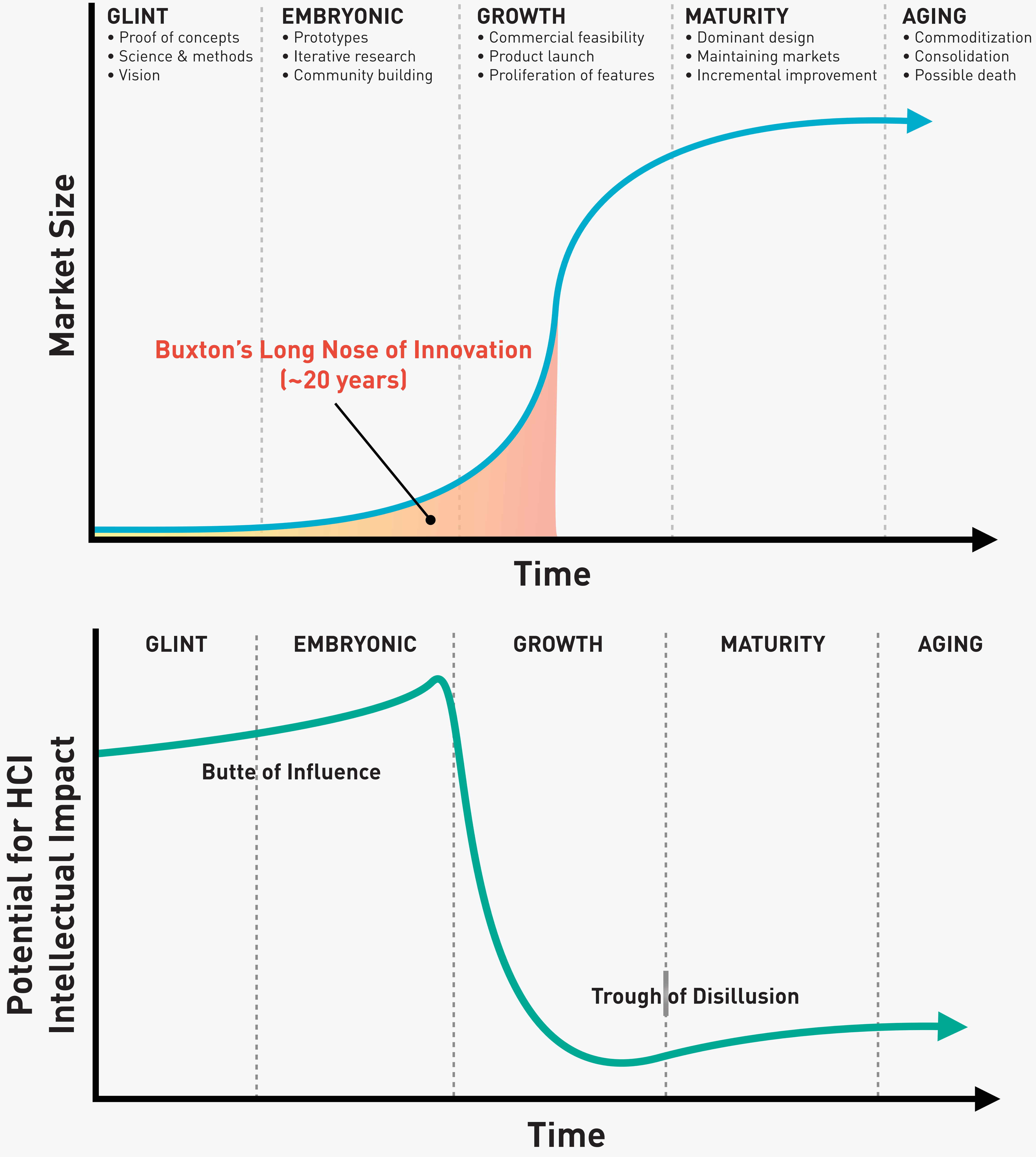
 Full video and description: thedemo.org



Harrison, C. (2018). The HCI innovator's dilemma. Interactions, 25(6), 26-33.



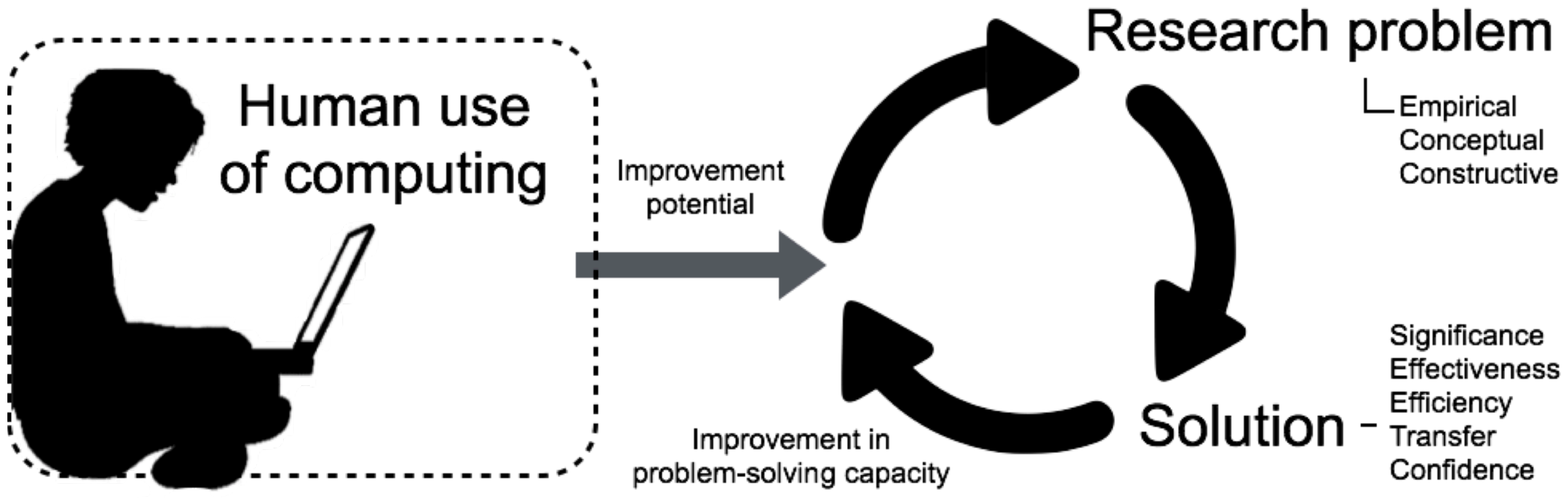
 Harrison, C. (2018). The HCI innovator's dilemma. Interactions, 25(6), 26-33.

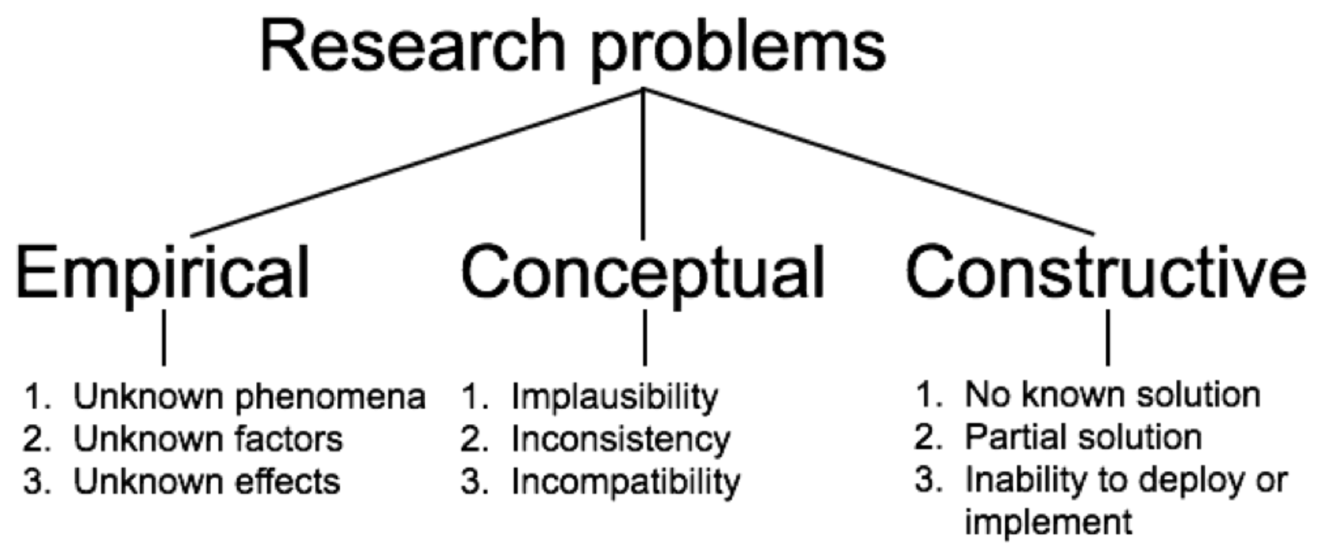


Harrison, C. (2018). The HCI innovator's dilemma. Interactions, 25(6), 26-33.

Many ways to validate constructive research

Ways to validate research in general





Criterion	**Evaluation Criteria**	**Heuristics for Refining Ideas**
Significance	Number of stakeholders involved; importance of the improvement for stakeholders; costs incurred when the improvement is *not* achieved	Target a different stakeholder group or a larger number of stakeholders; aim at a greater improvement over the present baseline; report on direct comparisons against baseline solutions
Effectiveness	Capture the essential aspects of the problem; match between evaluation metrics and priorities	Use multiple evaluation criteria and richer evaluation contexts; validate evaluation criteria; address unnoticed real-world difficulties
Efficiency	How much effort or resources it takes to create or deploy the solution; scalability; size	Develop tools for practitioners; share datasets and code; reduce price/cost
Transfer	Number of users, tasks, and contexts for which the solution can be applied; qualitatively new contexts wherein the solution can be applied	Identify and target new user groups, contexts, or tasks; demonstrate broad-based generalisability
Confidence	Empirical validity; reliability; replicability; reproducibility; robustness	Replicate the result in different contexts; report on different metrics for judging validity and reliability; allow reanalysis

Oulasvirta, A., & Hornbæk, K. (2016, May). HCI research as problem-solving. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4956-4967).

TYPE 1 - DEMONSTRATION

- **INDIVIDUAL INSTANCES**
  - NOVEL EXAMPLES
  - REPLICATED EXAMPLES
- **COLLECTIONS**
  - CASE STUDIES
  - DESIGN SPACES
- **GOING BEYOND DESCRIPTIONS**
  - HOW TO SCENARIOS

TYPE 2 - USAGE

- **WAYS TO CONDUCT USAGE STUDIES**
  - USABILITY STUDIES
  - A/B COMPARISON
  - WALKTHROUGH
  - OBSERVATION
  - TAKE-HOME STUDIES
- **ELICITING USER FEEDBACK**
  - LIKERT SCALES
  - INTERVIEWS

TYPE 3 - PERFORMANCE

- BENCHMARK THRESHOLD
- BENCHMARK COMPARISON

TYPE 4 - HEURISTICS

- CHECKLISTS
- DISCUSSIONS
- TARGETING

Ledo et al. (2018). Evaluation strategies for HCI toolkit research. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

MIT Tangible Media Group [Video]

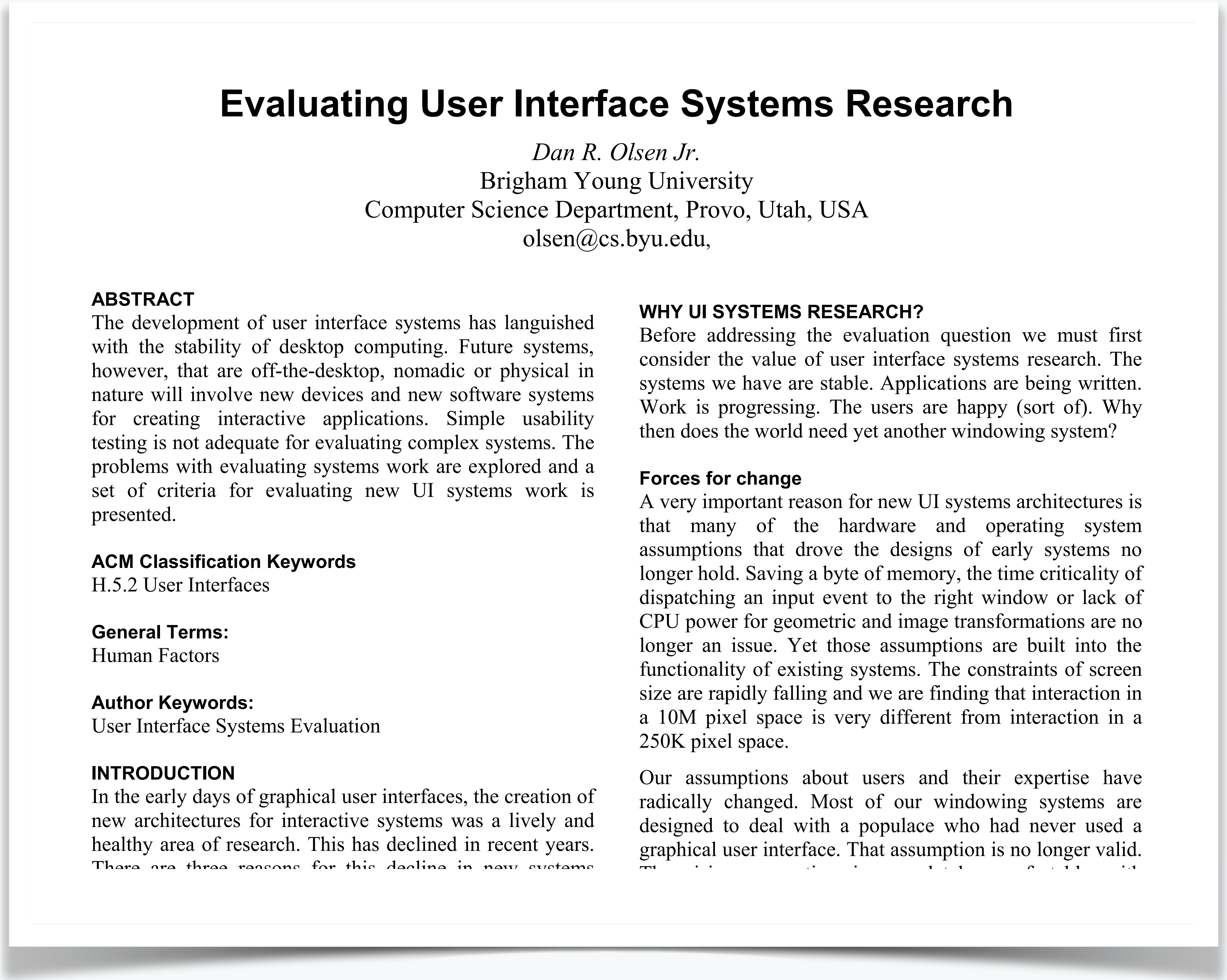
Sources of arguments for your systems

Olsen's concepts, e.g., **importance, generality, reducing viscosity, expressivity, combination.**

- Also recommended in UIST conference [guide for authors]

Formative validation with Green & Blackwell's **Cognitive Dimensions Framework:**

- Abstraction
- Hidden dependencies
- Premature commitment
- Secondary notation
- Viscosity
- Visibility


Evaluating User Interface Systems Research

Dan R. Olsen Jr.
Brigham Young University
Computer Science Department, Provo, Utah, USA
olsen@cs.byu.edu,

ABSTRACT
The development of user interface systems has languished with the stability of desktop computing. Future systems, however, that are off-the-desktop, nomadic or physical in nature will involve new devices and new software systems for creating interactive applications. Simple usability testing is not adequate for evaluating complex systems. The problems with evaluating systems work are explored and a set of criteria for evaluating new UI systems work is presented.

ACM Classification Keywords
H.5.2 User Interfaces

General Terms:
Human Factors

Author Keywords:
User Interface Systems Evaluation

INTRODUCTION
In the early days of graphical user interfaces, the creation of new architectures for interactive systems was a lively and healthy area of research. This has declined in recent years.

WHY UI SYSTEMS RESEARCH?
Before addressing the evaluation question we must first consider the value of user interface systems research. The systems we have are stable. Applications are being written. Work is progressing. The users are happy (sort of). Why then does the world need yet another windowing system?

Forces for change
A very important reason for new UI systems architectures is that many of the hardware and operating system assumptions that drove the designs of early systems no longer hold. Saving a byte of memory, the time criticality of dispatching an input event to the right window or lack of CPU power for geometric and image transformations are no longer an issue. Yet those assumptions are built into the functionality of existing systems. The constraints of screen size are rapidly falling and we are finding that interaction in a 10M pixel space is very different from interaction in a 250K pixel space.

Our assumptions about users and their expertise have radically changed. Most of our windowing systems are designed to deal with a populace who had never used a graphical user interface. That assumption is no longer valid.

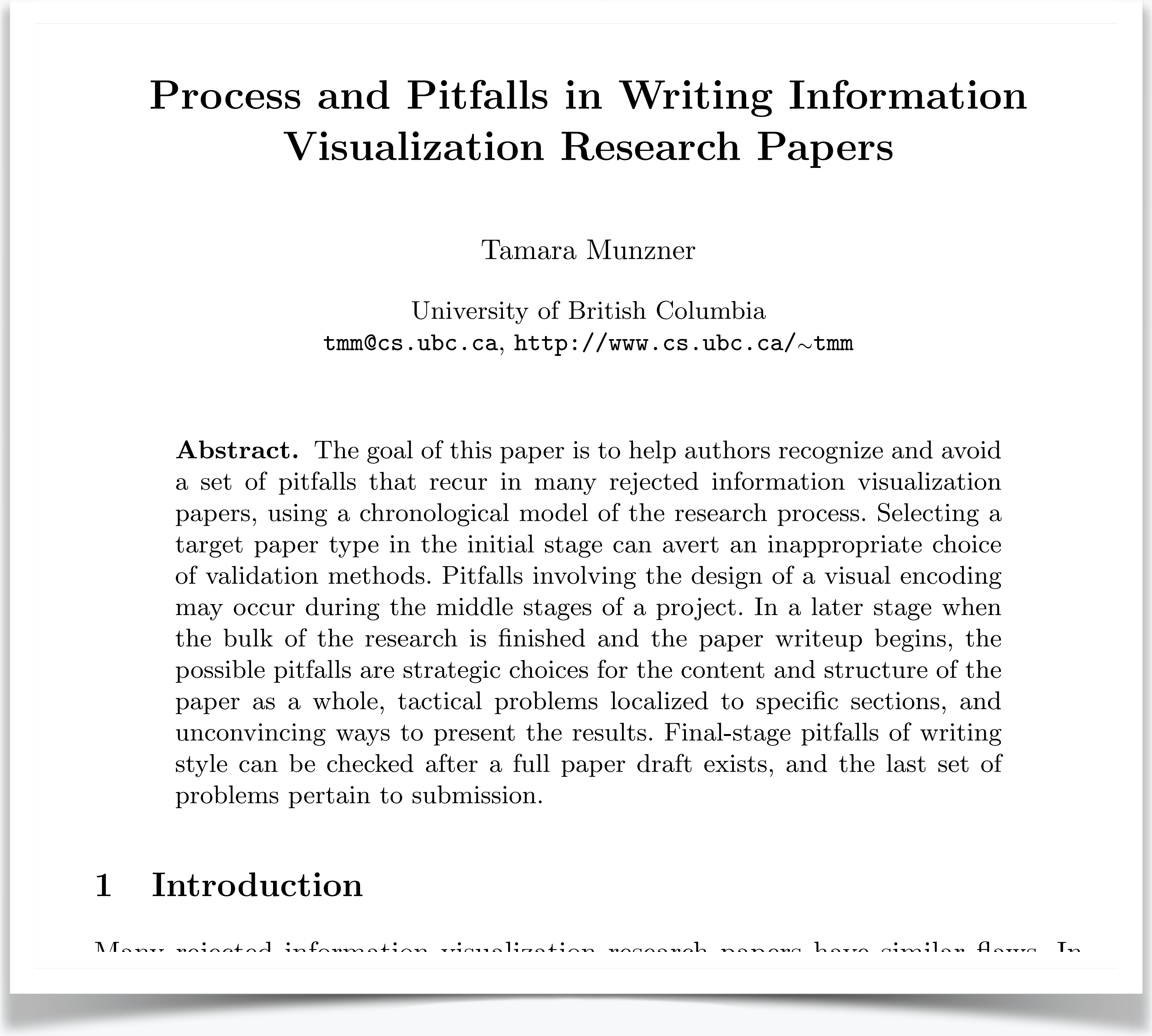
Olsen Jr, D. R. (2007). [Evaluating user interface systems research.] In Proceedings of the 20th annual ACM symposium on User interface software and technology (pp. 251-258).
Green, T., & Blackwell, A. (1998). [Cognitive dimensions of information artefacts: a tutorial]. In BCS HCI conference (Vol. 98, pp. 1-75). Sheffield, UK: Springer.

Guide for choosing validation approaches

Tamara Munzner (from the Information Visualization research, but broadly applicable)

Discuss pitfalls and how to prevent them

- All that coding means I deserve a systems paper
- Hammer in search of nail
- Dense as plutonium
- …



Process and Pitfalls in Writing Information Visualization Research Papers

Tamara Munzner

University of British Columbia
tmm@cs.ubc.ca, http://www.cs.ubc.ca/~tmm

Abstract. The goal of this paper is to help authors recognize and avoid a set of pitfalls that recur in many rejected information visualization papers, using a chronological model of the research process. Selecting a target paper type in the initial stage can avert an inappropriate choice of validation methods. Pitfalls involving the design of a visual encoding may occur during the middle stages of a project. In a later stage when the bulk of the research is finished and the paper writeup begins, the possible pitfalls are strategic choices for the content and structure of the paper as a whole, tactical problems localized to specific sections, and unconvincing ways to present the results. Final-stage pitfalls of writing style can be checked after a full paper draft exists, and the last set of problems pertain to submission.

1 Introduction

Munzner, T. (2008). Process and pitfalls in writing information visualization research papers. In *Information visualization: human-centered issues and perspectives* (pp. 134-153). Berlin, Heidelberg: Springer Berlin Heidelberg.

Exercise: Brainstorm ways to validate your research **Miro**

Take 5 minutes to think of 1–2 **validation strategies** that could be applied to your research problem. Add them to the Miro board.

~~Together with a person next to you, take 10 minutes per person to discuss your choice(s) of validation techniques:~~

- ~~Are these validation suitable for the research contribution?~~
- ~~How "useful" are these validations?~~
- ~~What other ways could be used to validate?~~

Validation strategies

TYPE 1 - DEMONSTRATION

- INDIVIDUAL INSTANCES
  - NOVEL EXAMPLES
  - REPLICATED EXAMPLES
- COLLECTIONS
  - CASE STUDIES
  - DESIGN SPACES
- GOING BEYOND DESCRIPTIONS
  - HOW TO SCENARIOS

TYPE 2 - USAGE

- WAYS TO CONDUCT USAGE STUDIES
  - USABILITY STUDIES
  - A/B COMPARISON
  - WALKTHROUGH
  - OBSERVATION
  - TAKE-HOME STUDIES
- ELICITING USER FEEDBACK
  - LIKERT SCALES
  - INTERVIEWS

TYPE 3 - PERFORMANCE

- BENCHMARK THRESHOLD
- BENCHMARK COMPARISON

TYPE 4 - HEURISTICS

- CHECKLISTS
- DISCUSSIONS
- TARGETING

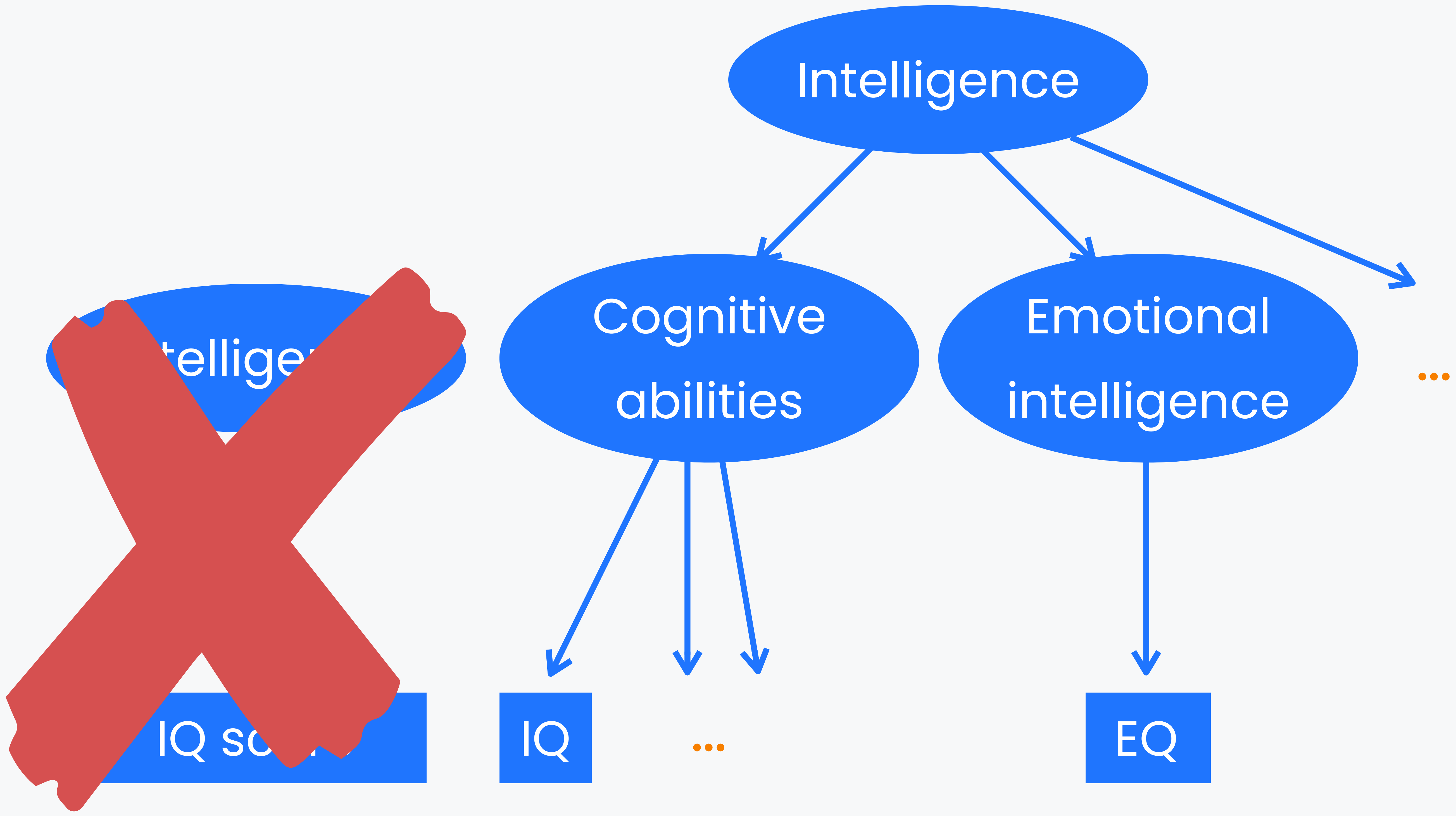
Tools for thinking about empirical research

Constructs vs. Operational definition

Construct: Theoretical and latent concepts that (we expected to) help explain and predict a phenomenon

Operational definition: A specification of a procedure to manipulate or measure an external, observable the phenomenon

Construct validity: How much the operationalized measurement correspond to the construct of interest

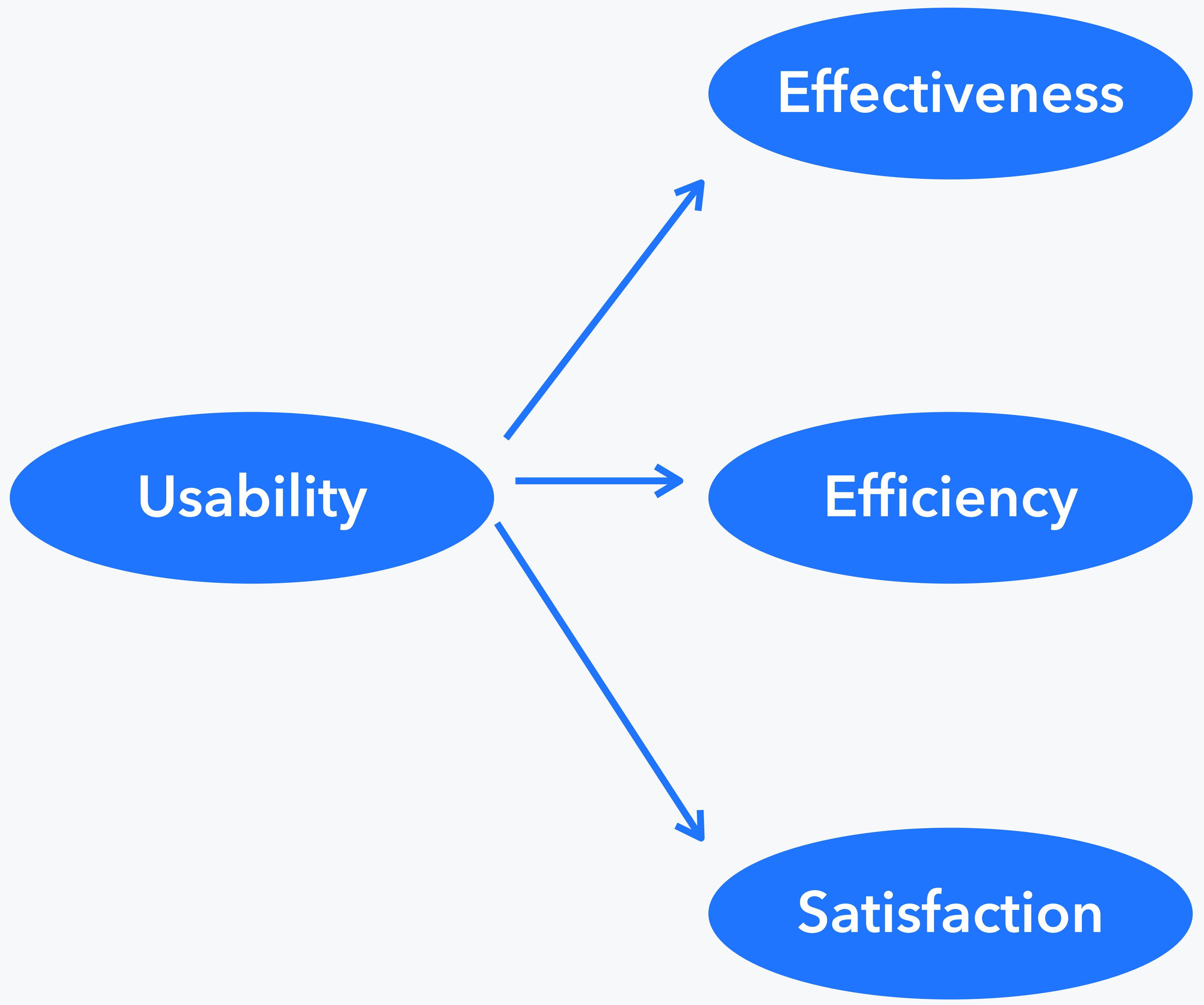


A construct may have multiple facets

Effectiveness: accuracy and completeness with which users achieve specified goals.

Efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals

Satisfaction: freedom from discomfort, and positive attitudes towards the use of the product



 ISO 9241 standard for usability

A construct may be operationalized in multiple ways

- Usability
  - Effectiveness
    - Accuracy: Number of errors, ratio of errors vs. success
    - Completeness: Number of tasks solved
    - Quality: Experts’ score of the outcome of the interactions
    - …
  - Efficiency
    - Time: Task completion time
    - Input rate: Keystrokes per minute
    - Mental effort: Users’ rating of their mental effort, users’ performance in their secondary task
    - …
  - Satisfaction
    - Preference: Rank preferred interface
    - Ease of use: Users’ rating
    - Perception of outcome: Users’ rating on sense of success
    - …

For more examples, see Hornbæk, K. (2006). [Current practice in measuring usability: Challenges to usability studies and research.]() International journal of human-computer studies, 64(2), 79-102.

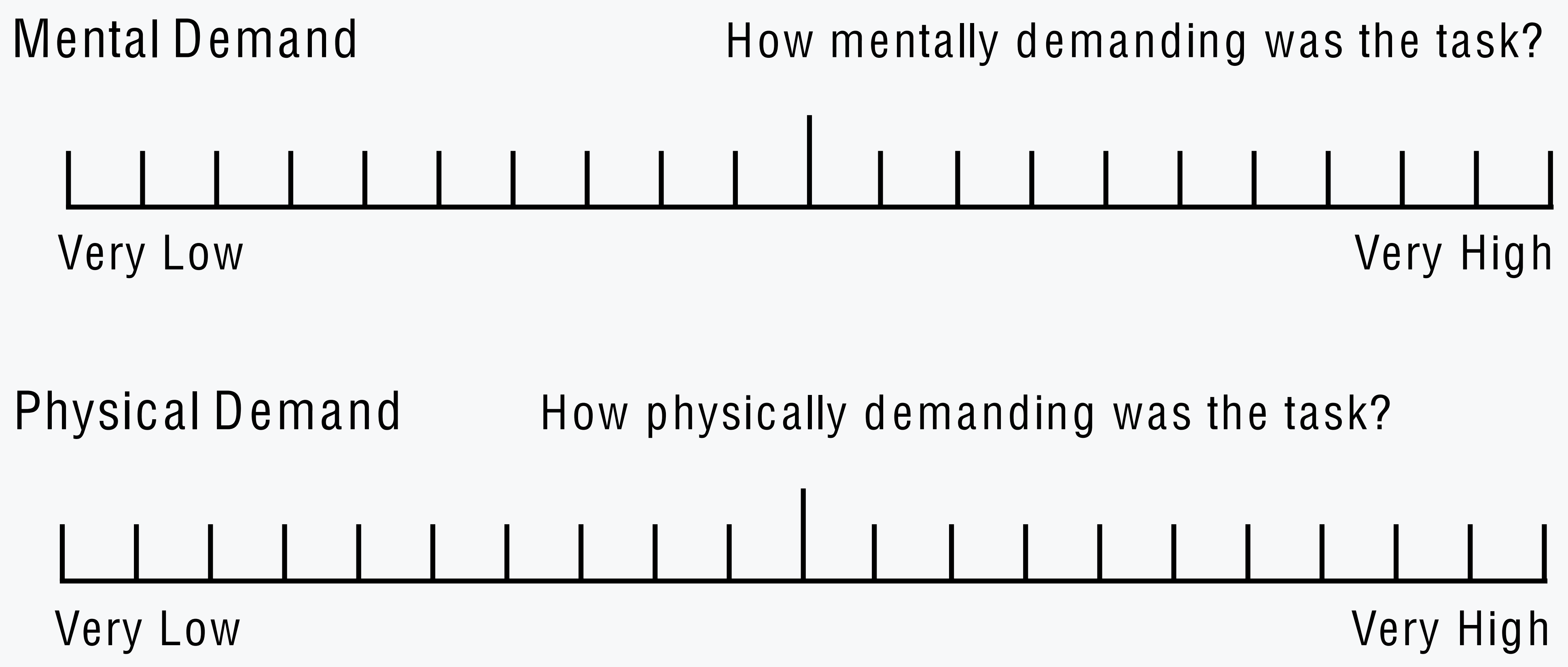
Constructs can be measured or manipulated

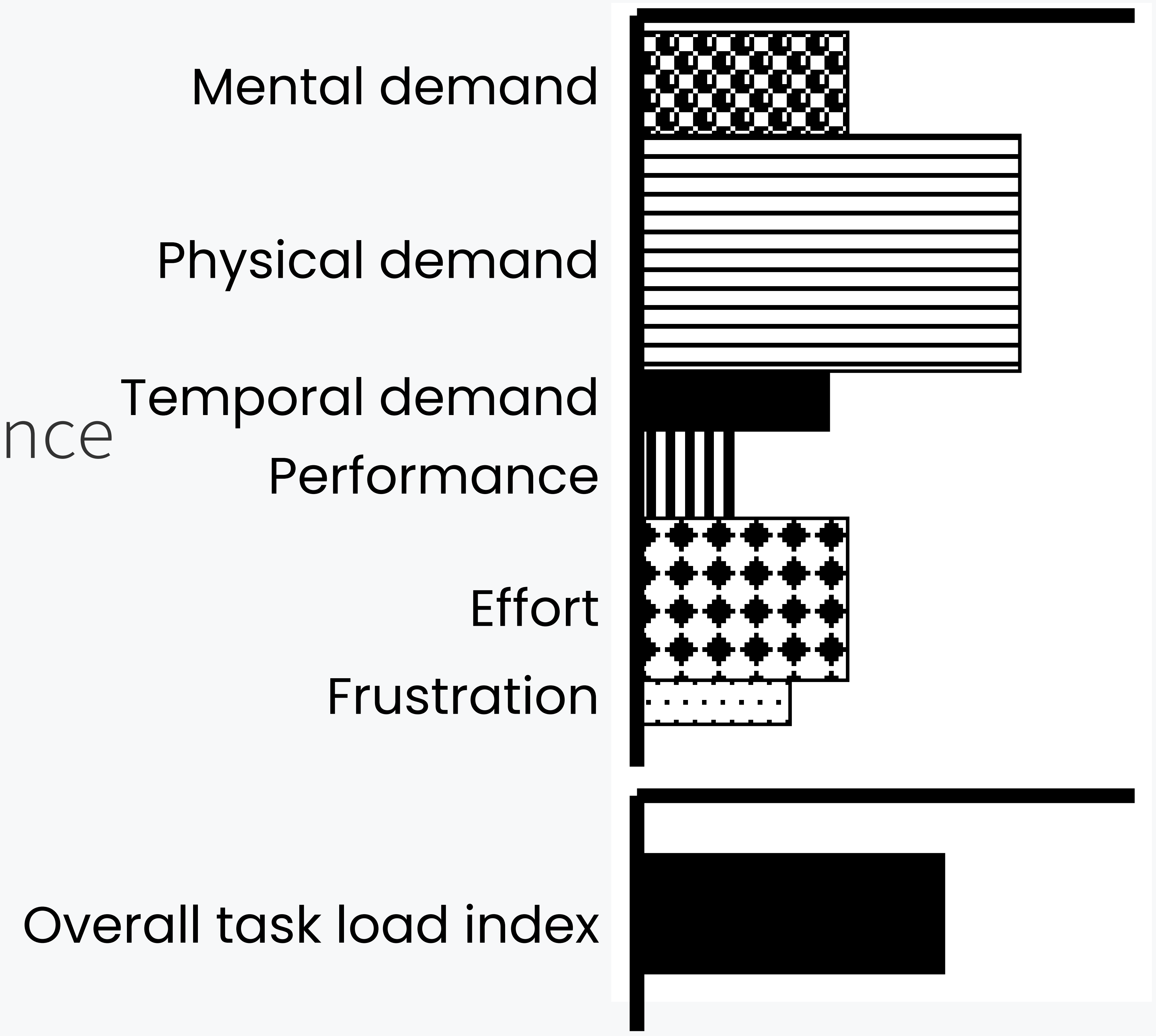
Example: measuring mental workload

NASA Task Load Index (NASA-TLX)

- Rating on a 100-point range
- Pairwise comparison of subscales' importance

Example questions:





Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Constructs can be measured or manipulated

Example: manipulating mental workload

Purpose: Determine how much eye tracking could be used to estimate mental workload

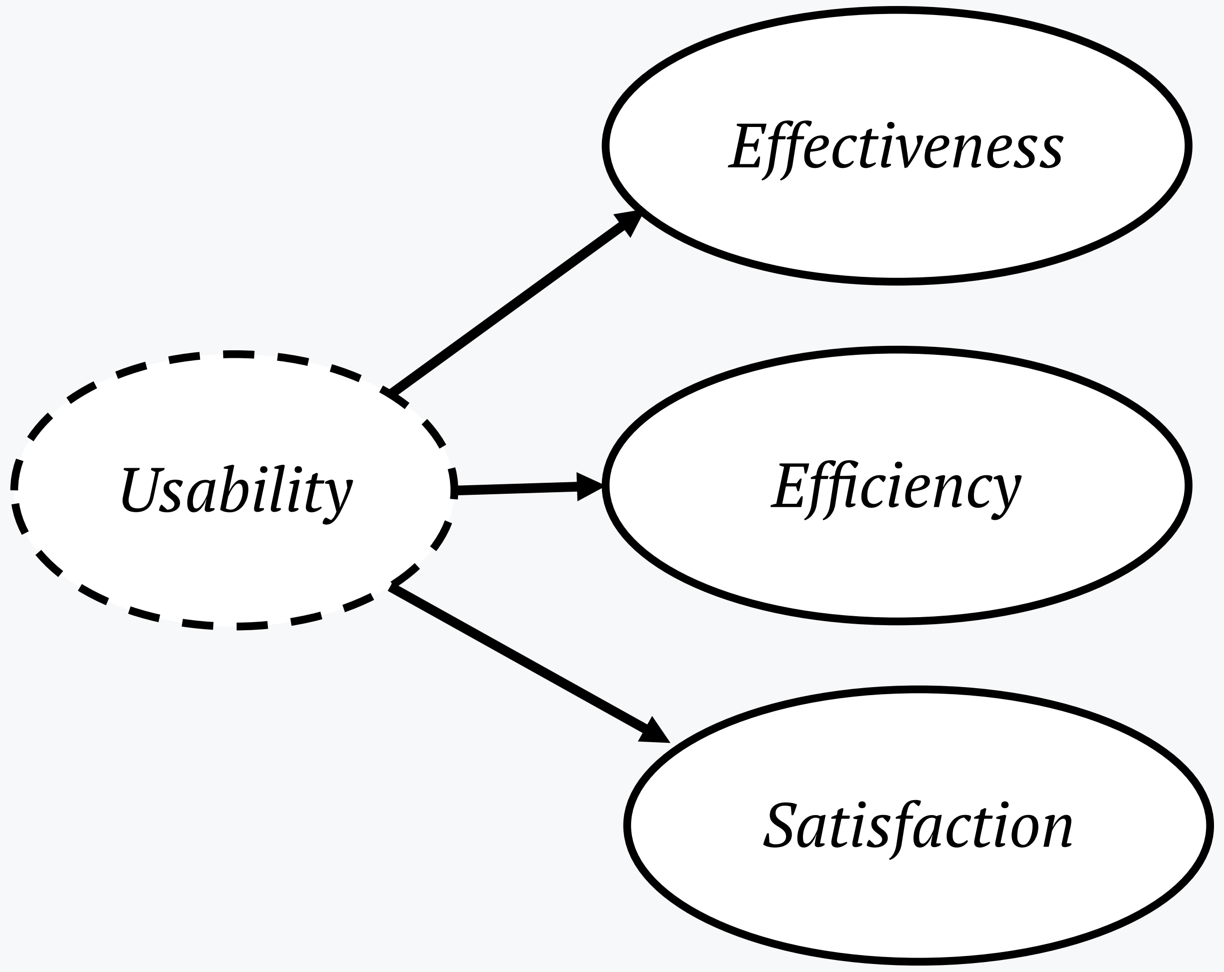
Manipulation: Asking the driver to remember single digit number and repeat back verbally

- Immediately
- After another number was presented
- After another two numbers were presented

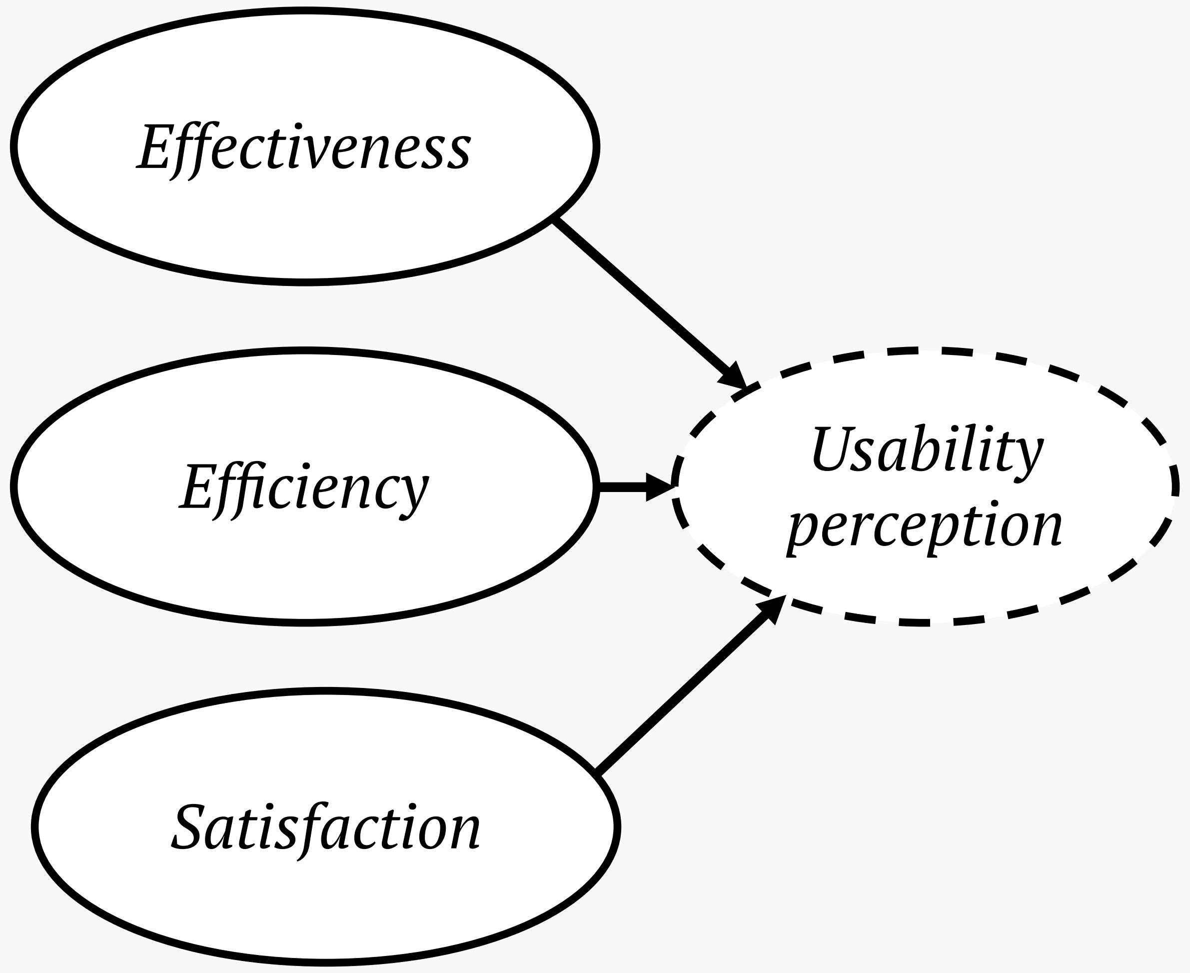
Fridman, L., Reimer, B., Mehler, B., & Freeman, W. T. (2018, April). Cognitive load estimation in the wild. In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1-9).

Different theoretical assumptions

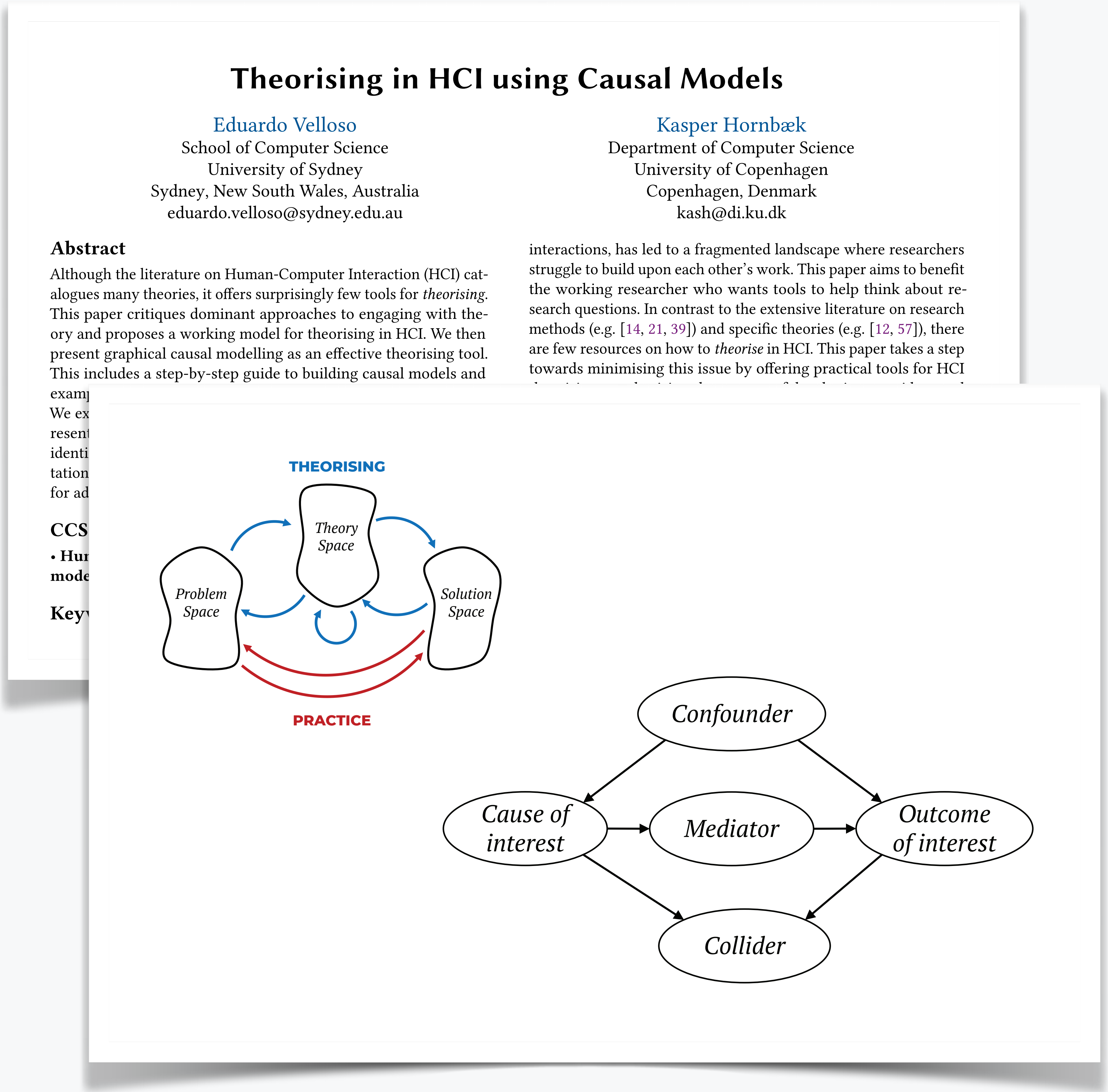
E,E,S estimates the latent usability construct



E,E,S predicts the latent usability construct



Section 3 in the paper includes good explanation of these term and concrete examples from HCI



 Velloso, E., & Hornbæk, K. (2025). Theorising in HCI using Causal Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Exercise: Think about constructs and operationalization **Miro**

Take 5 minutes to think of **one construct** and **two ways to operationalize** it. Write your thoughts on the Miro board.

If your research doesn't use empirical validation, do this task as a thought experiment

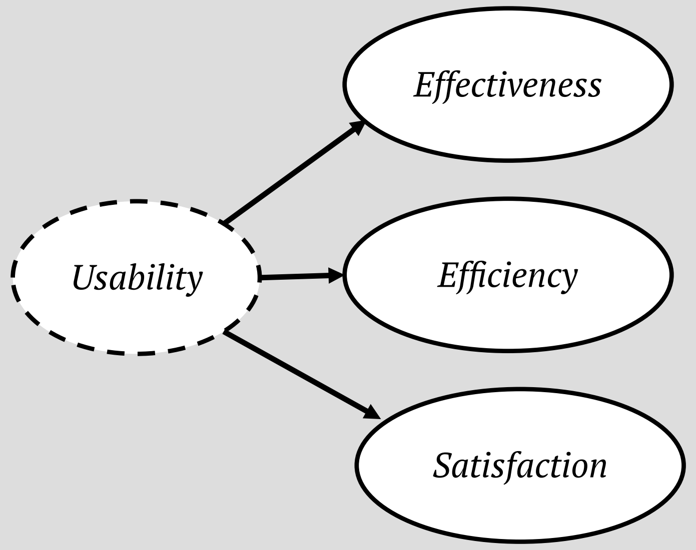
~~Together with a person next to you, take 10 minutes per person to discuss:~~

- ~~Are the operationalization correspond to the construct?~~
- ~~What are other ways to operationalize?~~

Construct: Theoretical and latent concepts that (we expected to) help explain and predict a phenomenon

Operational definition: A specification of a procedure to manipulate or measure an external, observable the phenomenon

Examples:



Effectiveness:

- Accuracy: Number of errors, ratio of errors vs. success
- Number of tasks solved

Efficiency:

- Task completion time
- Mental effort rating

Types of research by **variable relationships**

Descriptive research: constructing an accurate description of what happened

"Y happened"

- Observation: Among 10 teens who play a specific game, 8 can touch-type. Among 12 teens who did not play, 2 can touch-type.

Relational research: identify relationships between multiple variables

"When Y happens, X also happens"

- Log number of gaming hours per week and measure their typing speed. Found a correlation that the higher gaming hours, the higher the typing speed

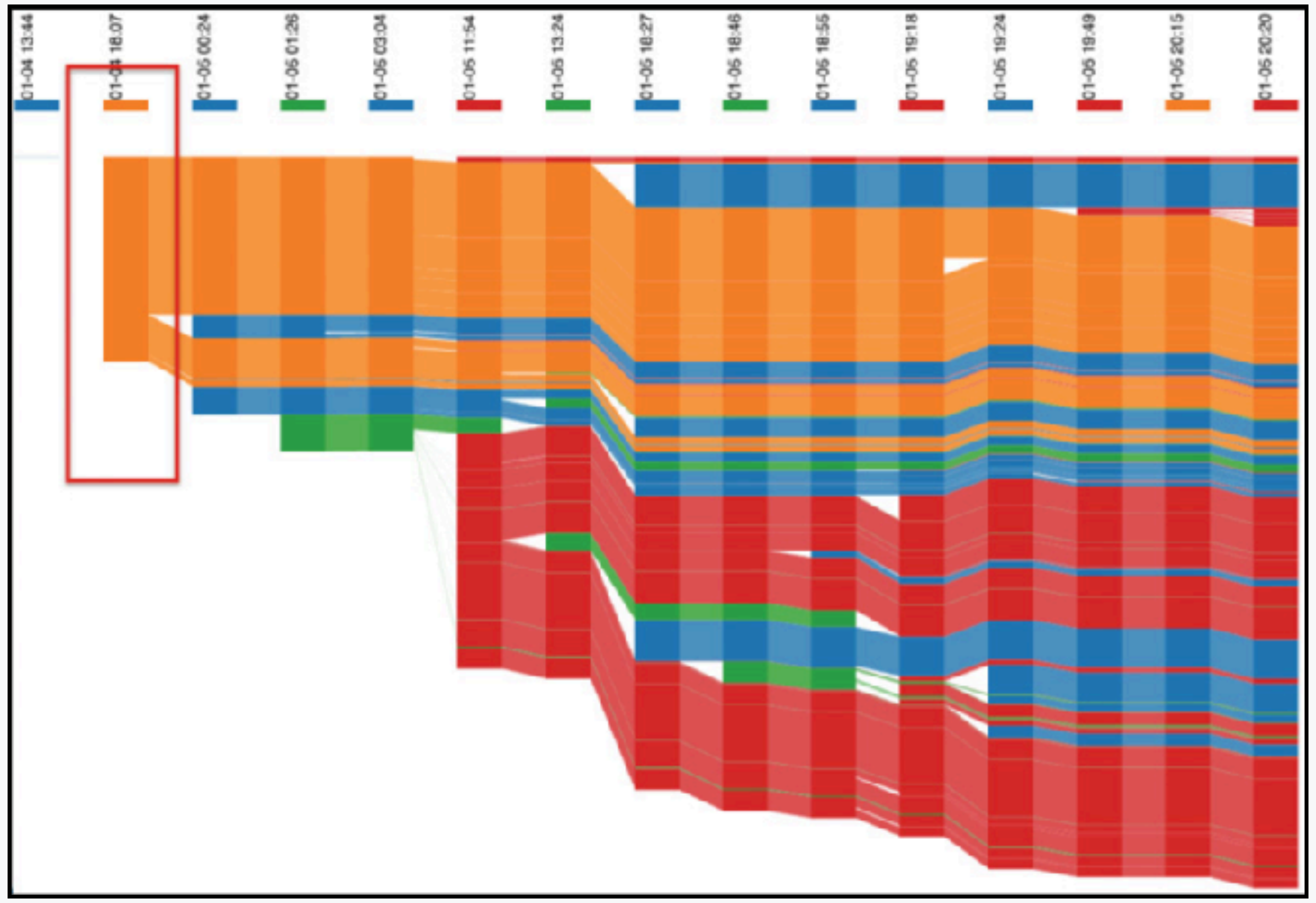
Experimental research: determine causal relationships between variables

"Y happens because of X"

- Randomly assign teens into 2 groups. One group is assigned to play the game a certain hours per week, the other not. After 3 months, the typing speed of the gaming group is higher than the non-gaming group.

Example: collaboration pattern on google doc

- Data: Interaction traces from 96 Google Docs from students' work in a semester
- Researchers group the traces into collaboration styles
- These styles are then associated with the writing quality rated by experts
- Some collaboration styles yielded higher writing quality than others



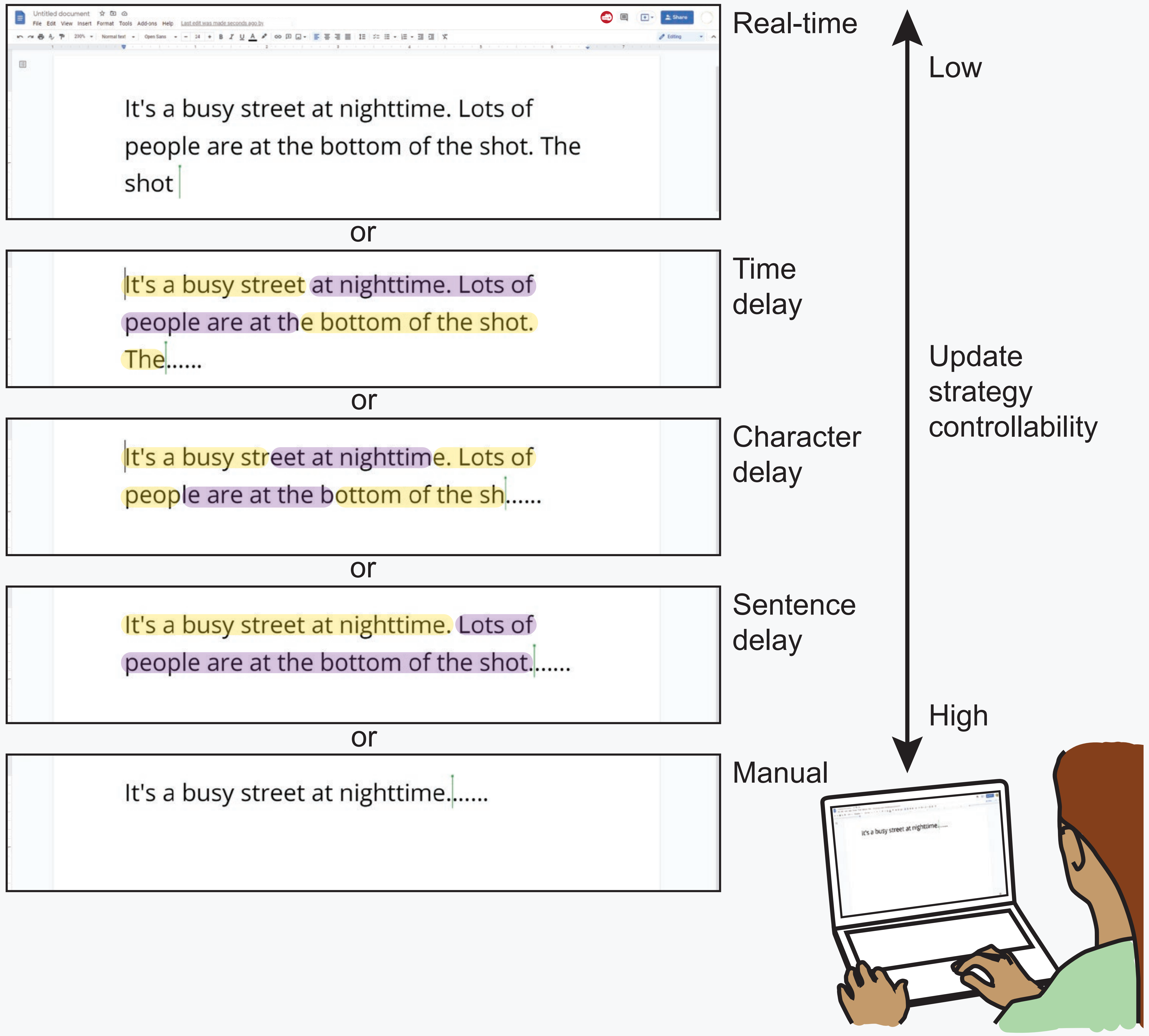
- Descriptive research: "Y happened"
- Relational research: "When Y happens, X also happens"
- Experimental research: "Y happens because of X"

Olson et al. (2017) [How People Write Together Now]

Example: Update intervals for editing shared documents

- Collaboration behavior on Google Docs
- Different update intervals are presented to the observers
- Observers rate their experience (e.g., ability to follow updates, naturalness)
- Results: Different strategies yielded different ratings

(b) Observer's view **for different update strategies**.



- Descriptive research: "Y happened"
- Relational research: "When Y happens, X also happens"
- Experimental research: "Y happens because of X"

Yeh et al. (2024) [The Effects of Update Interval and Reveal Method on Writer Comfort in Synchronized Shared-Editors]

Goal:
Increase second week retention

↓

Data:

"Only 30% of users can complete tasks using current navigation."

"When navigation was changed in a prior test, retention went down."

"It is becoming a UX best practice to NOT use a 'hamburger menu' navigation."

↓

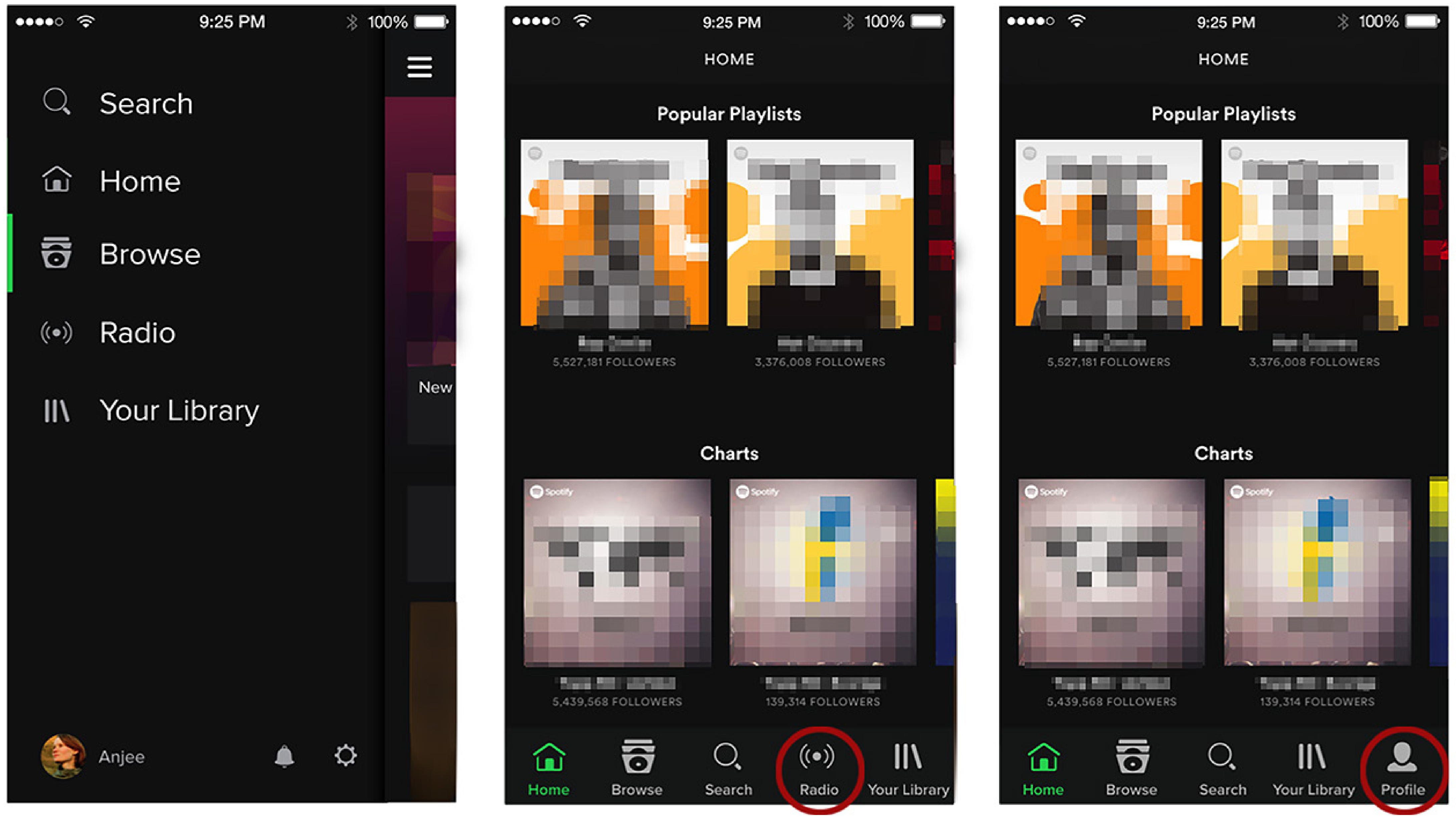
Problem/Opportunity Area:
Navigation

↓

Hypotheses

- clarify the value of Spotify (what Spotify offers) by simplifying the information architecture
- make it easier to discover features by making the navigation more prominent

Conditions: Three designs



Measurement: Second week retention rate

- Descriptive research: "Y happened"
- Relational research: "When Y happens, X also happens"
- Experimental research: "Y happens because of X"

King, R., Churchill, E. F., & Tan, C. (2017). Designing with data: Improving the user experience with A/B testing. O'Reilly Media, Inc.

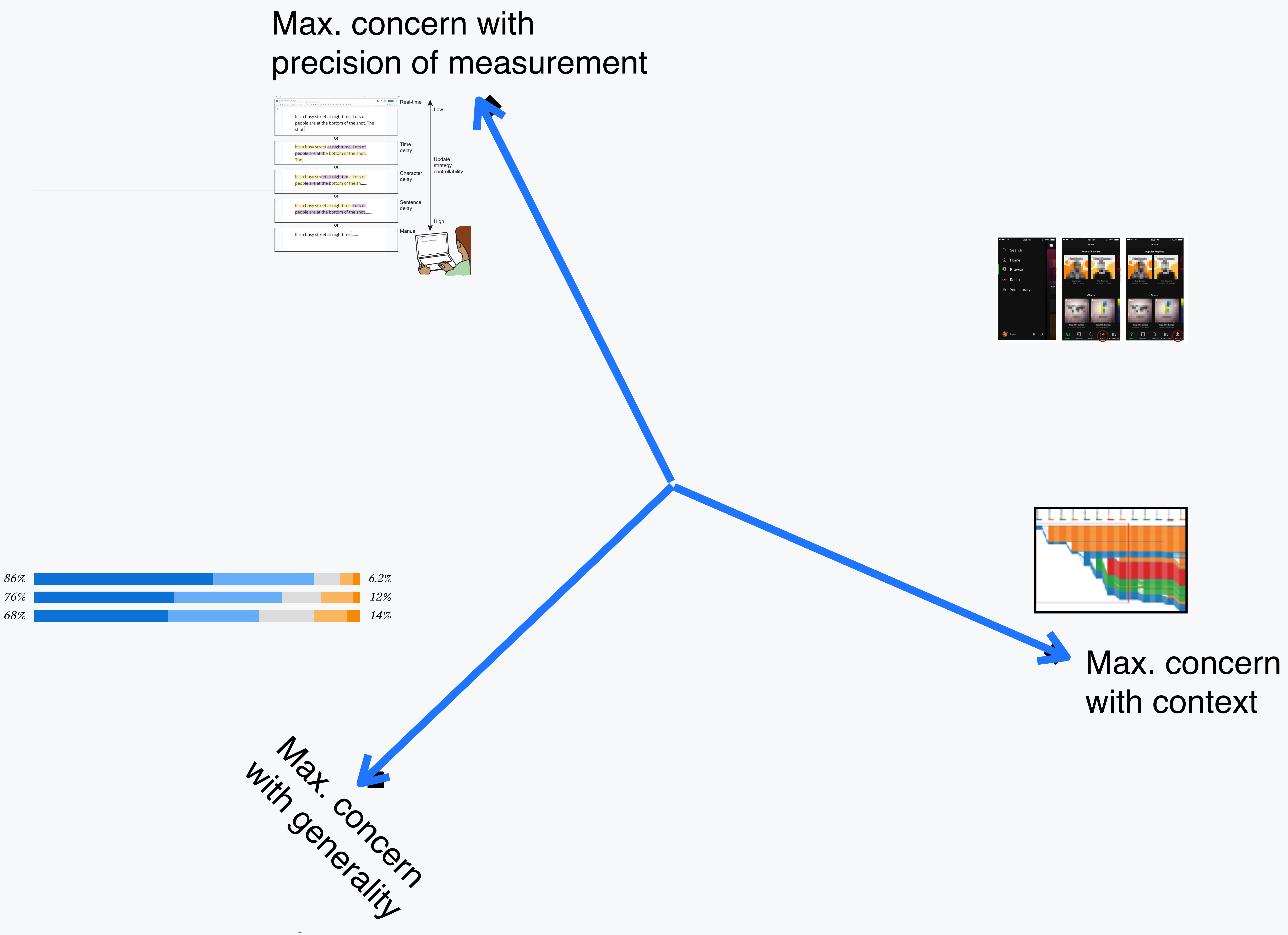
Example: Survey on AI usage in coding

	Motivation	Distribution	
	A. For using		
M1	To have an autocomplete or reduce the amount of keystrokes I make.	*86%*	*6.2%*
M2	To finish my programming tasks faster.	*76%*	*12%*
M3	To skip needing to go online to find specific code snippets, programming syntax, or API calls I'm aware of, but can't remember.	*68%*	*14%*
	B. For not using		
M6	Code generation tools write code that doesn't meet functional or non-functional (e.g., security, performance) requirements that I need.	*54%*	*34%*
M7	It's hard to control code generation tools to get code that I want.	*48%*	*36%*
M8	I spend too much time debugging or modifying code written by code generation tools.	*38%*	*45%*

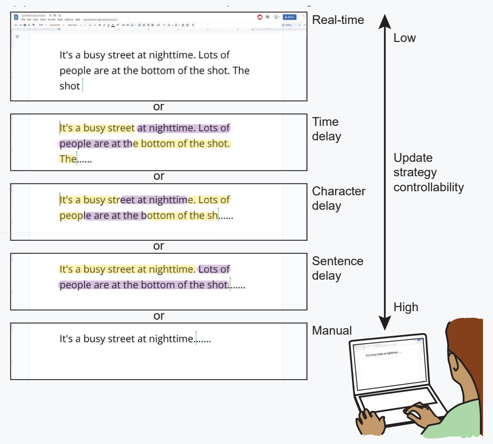
Very important · Important · Moderately important · Slightly important · Not important at all

- Descriptive research: "Y happened"
- Relational research: "When Y happens, X also happens"
- Experimental research: "Y happens because of X"

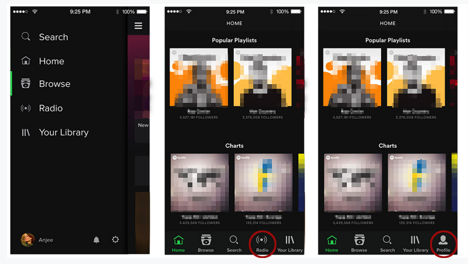
Liang, J. T., Yang, C., & Myers, B. A. (2024, February). A large-scale survey on the usability of ai programming assistants: Successes and challenges. In Proceedings of the 46th IEEE/ACM international conference on software engineering (pp. 1-13).

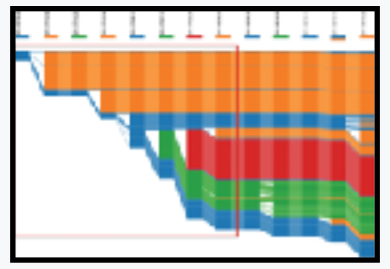

Max. concern with
precision of measurement
Real-time
Time delay
Character delay
Sentence delay
Manual
Low
Update strategy controllability
High
86%
76%
68%
6.2%
12%
14%
Max. concern
with context
Max. concern
with generality

Max. concern with
precision of measurement



Obtrusive
Research
Operations



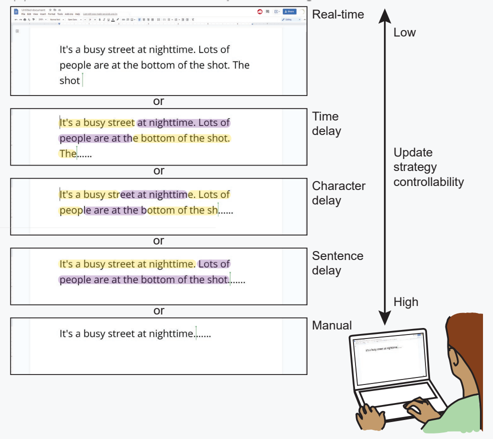


Max. concern
with context

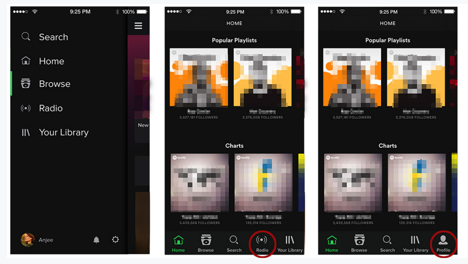
Unobtrusive
Research
Operations

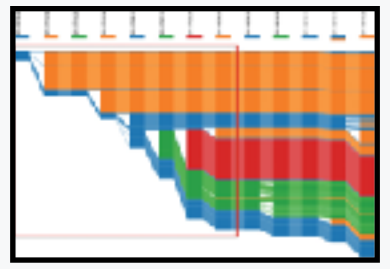
Max. concern
with generality

Max. concern with precision of measurement



Obtrusive Research Operations





Max. concern with context

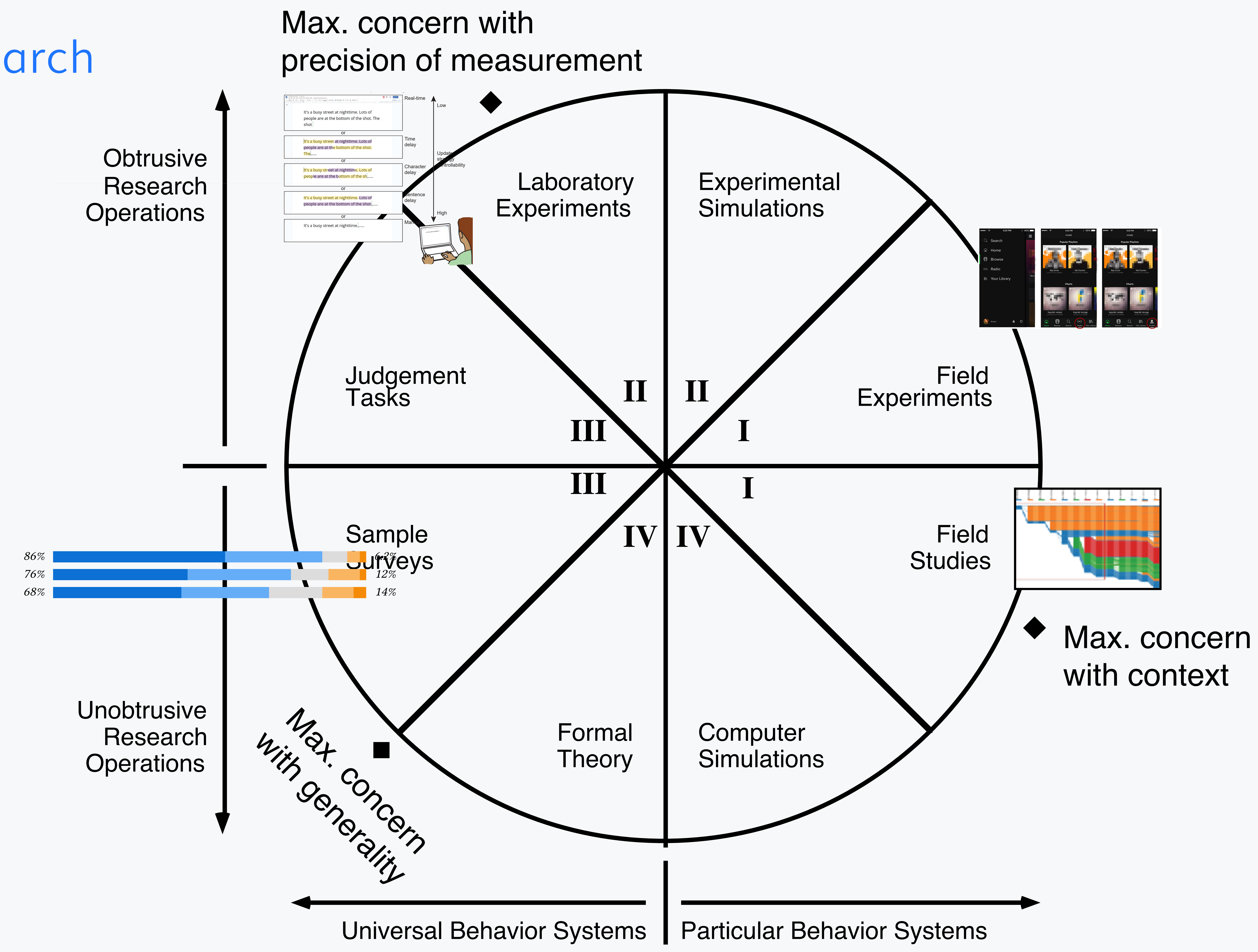
Unobtrusive Research Operations

Max. concern with generality

Universal Behavior Systems

Particular Behavior Systems

Strategies for Empirical research

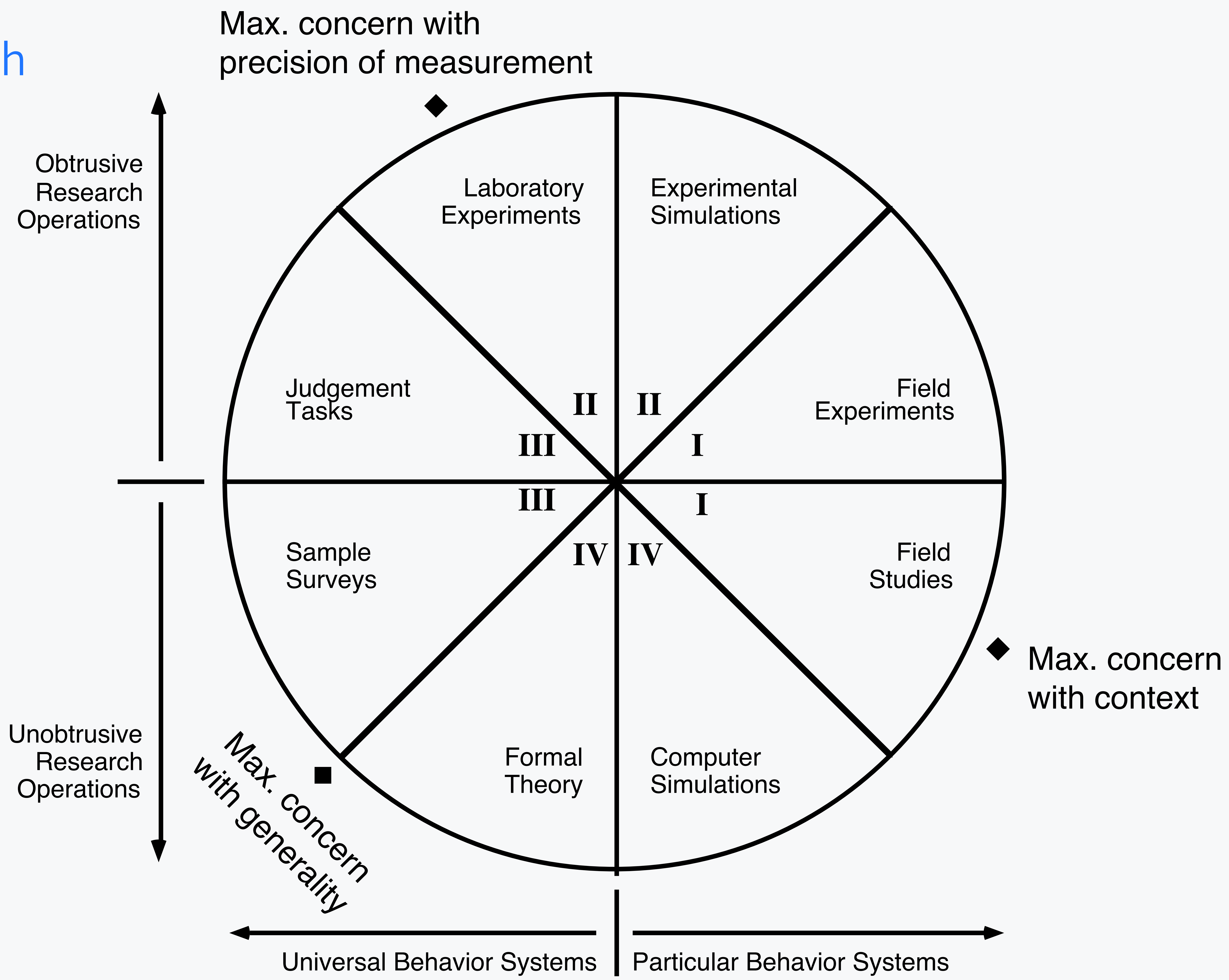


McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. American Behavioral Scientist, 25(2), 179-210.

McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In Readings in Human–Computer Interaction (pp. 152-169). Morgan Kaufmann.

Strategies for Empirical research

- No single method is perfect
- Use more than one research approach to address the same question and triangulate the findings
- Acknowledge limitations of your method and point to the next direction that you think needed the most



McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. American Behavioral Scientist, 25(2), 179-210.
McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In Readings in Human–Computer Interaction (pp. 152-169). Morgan Kaufmann.

Miro

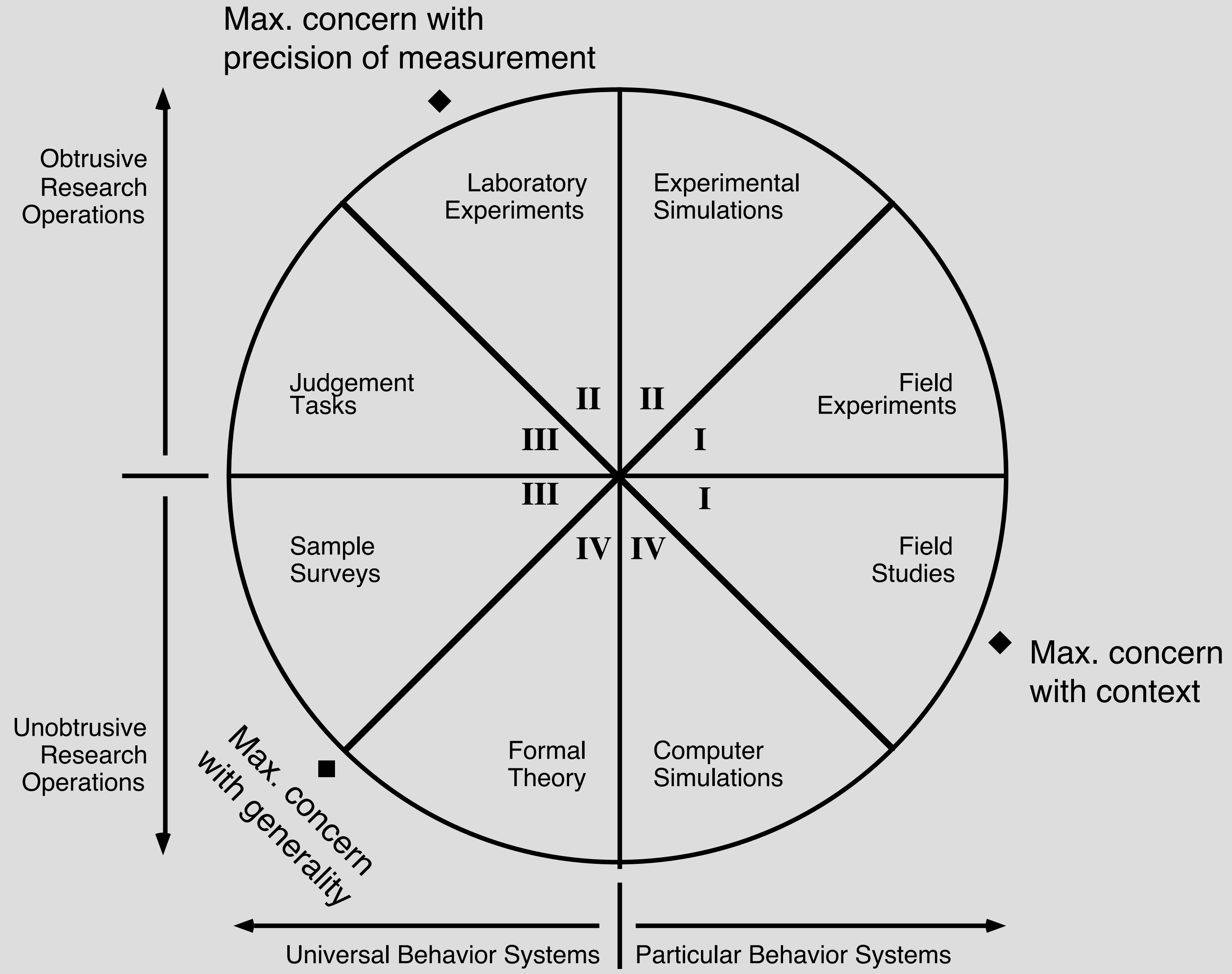
Take 10 minutes to think of **one variable-relationship** in your research. Select an **empirical strategy** to inspect this relationship.

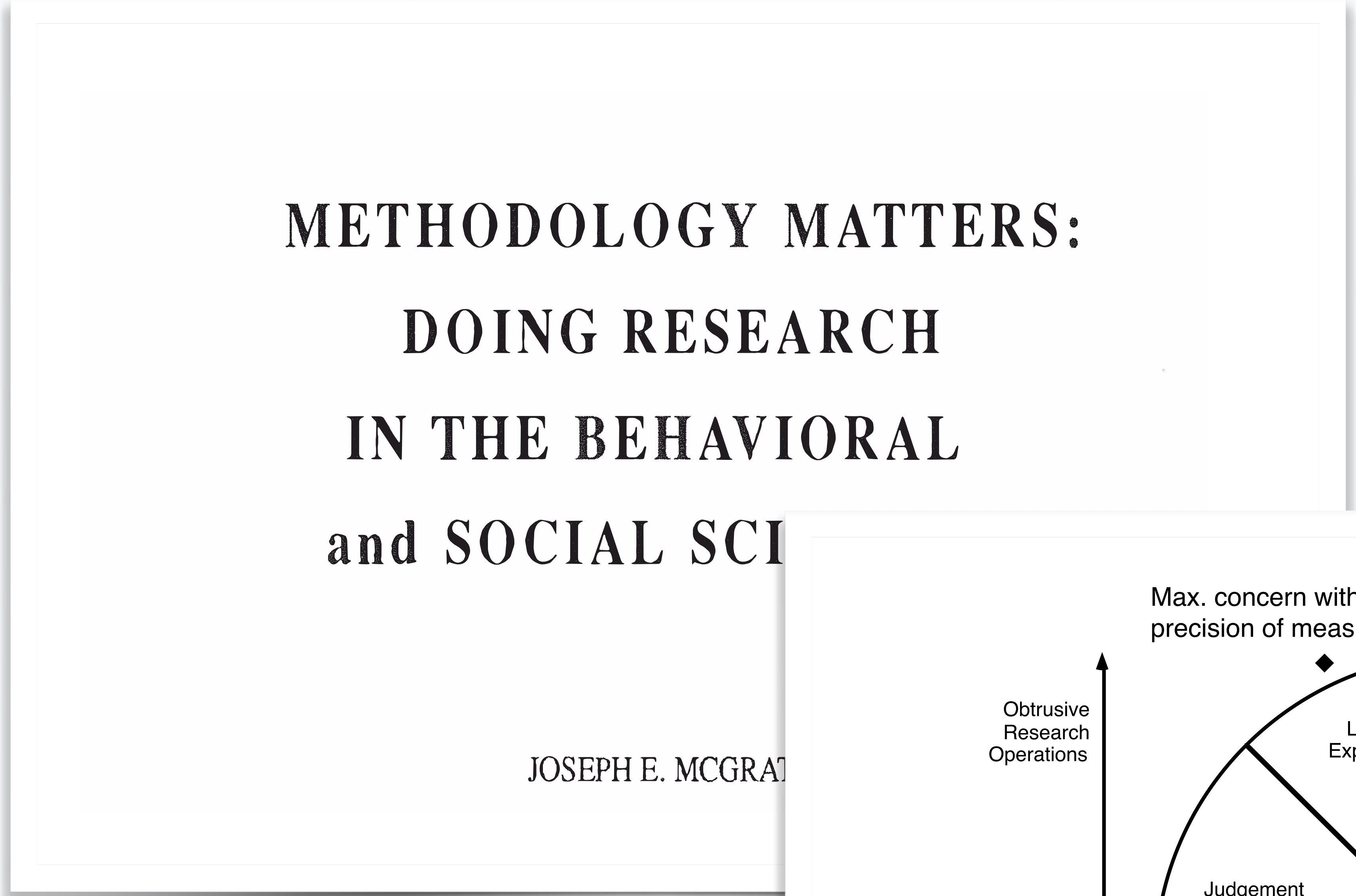
If your research doesn't use empirical validation, do this task as a thought experiment

~~Together with a person next to you, take 20 minutes per person to discuss:~~

- ~~Are the selected empirical strategy suitable to investigate the relationship?~~
- ~~What are strengths and weaknesses of this strategy? Are the weaknesses strong concern for this particular research problem?~~
- ~~What other strategies might be relevant?~~

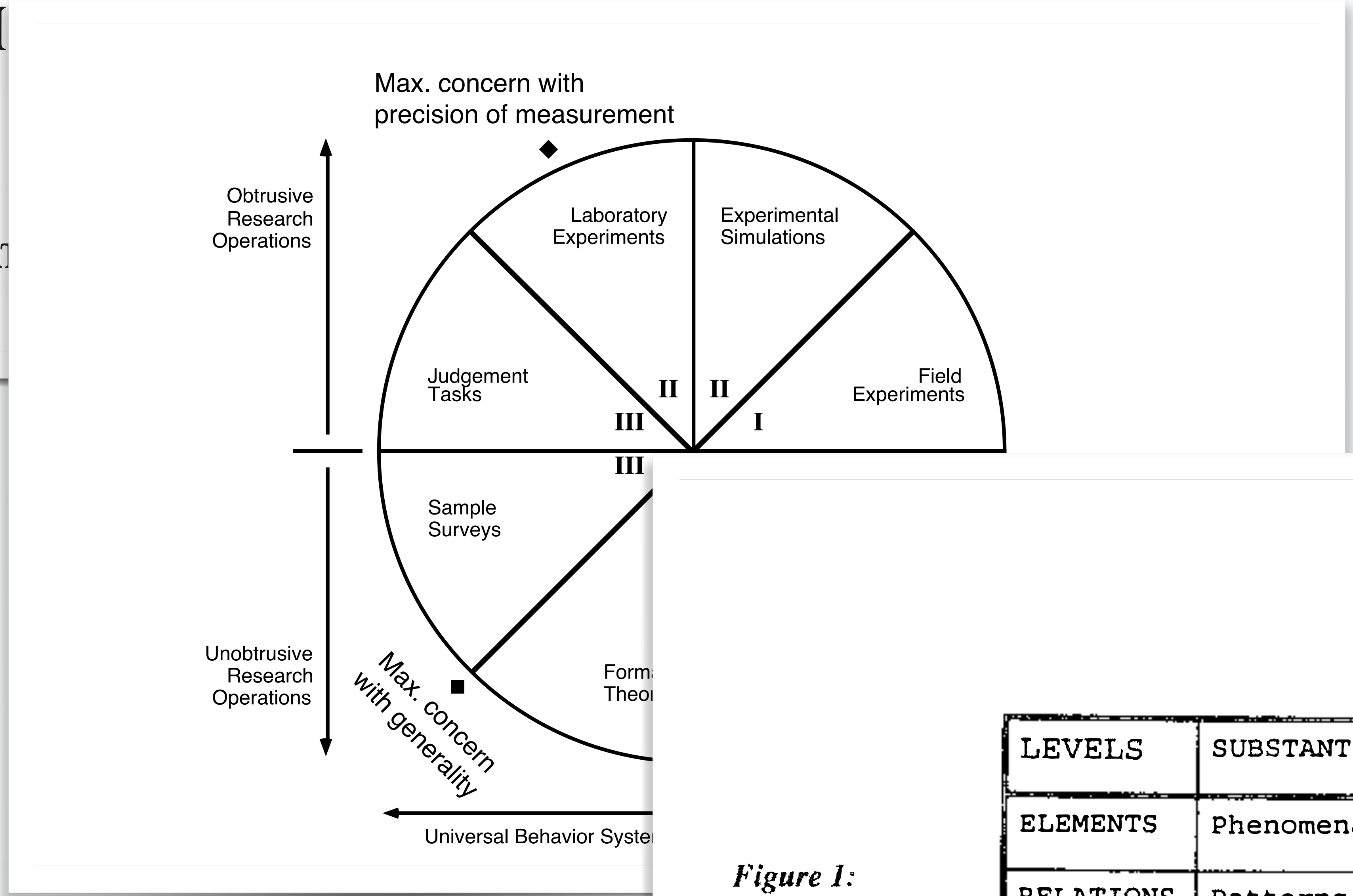
- **Descriptive research:** "Y happened"
- **Relational research:** "When Y happens, X also happens"
- **Experimental research:** "Y happens because of X"

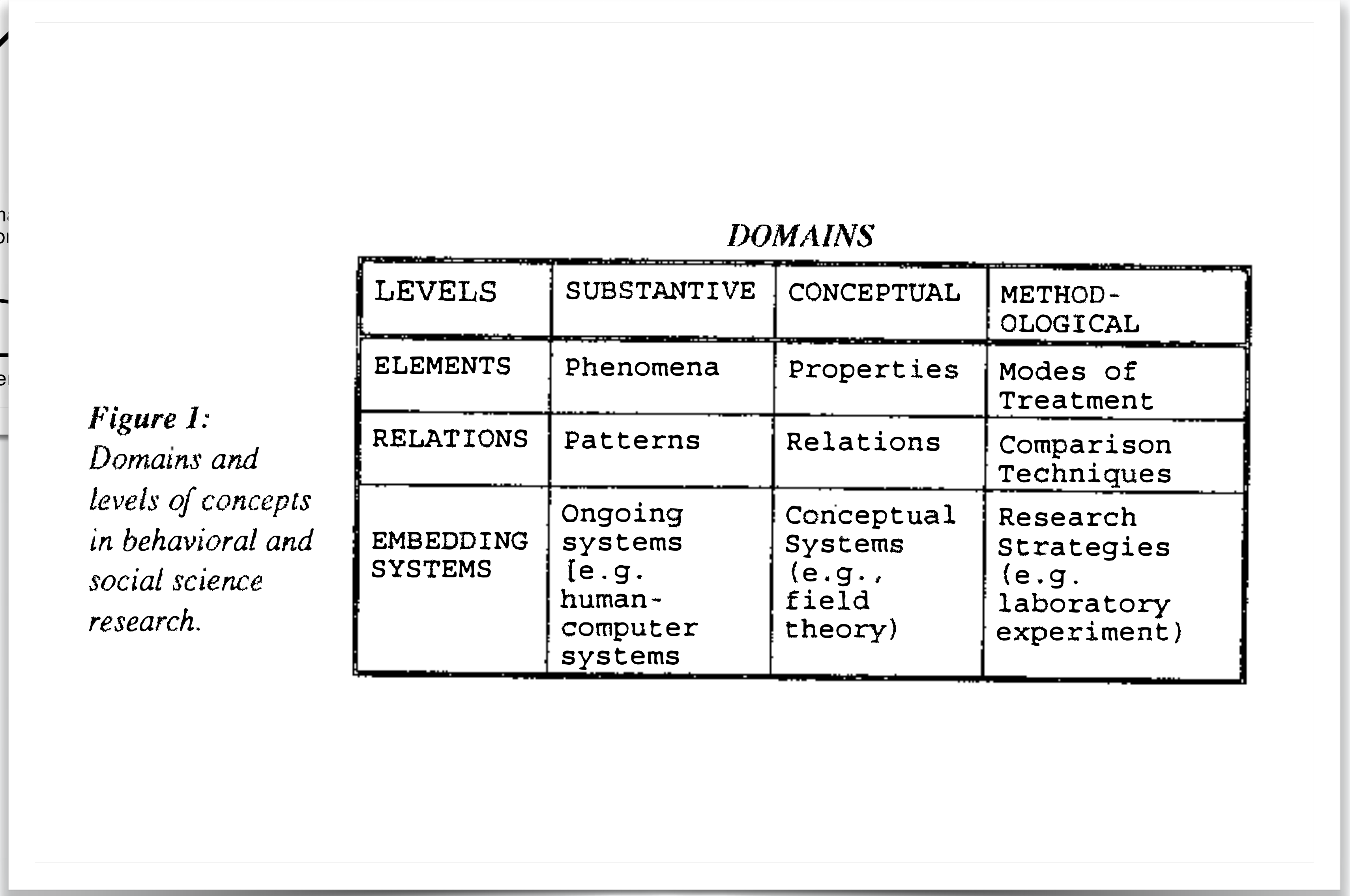




METHODOLOGY MATTERS:
DOING RESEARCH
IN THE BEHAVIORAL
and SOCIAL SCI

JOSEPH E. MCGRAT





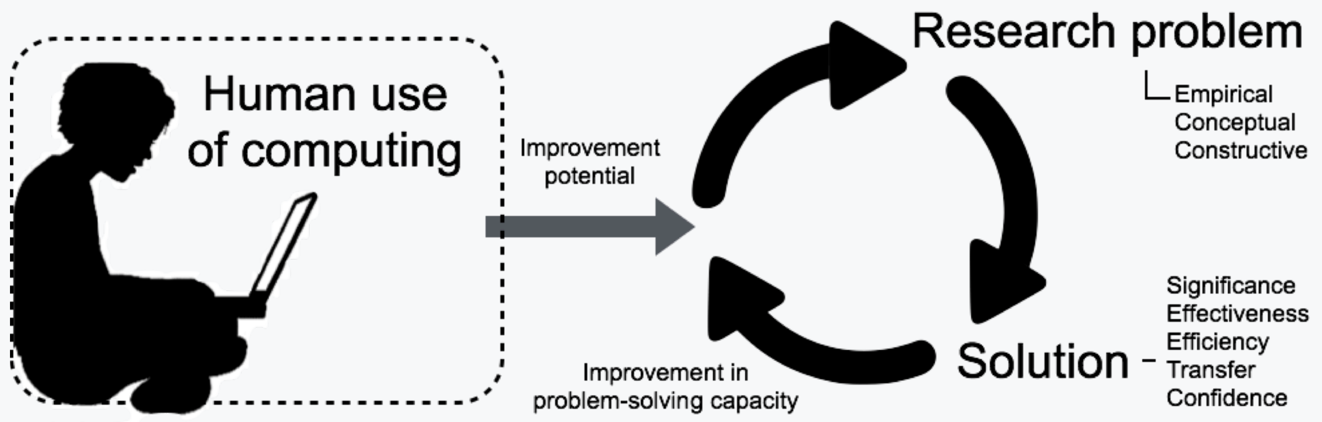
DOMAINS

LEVELS	SUBSTANTIVE	CONCEPTUAL	METHOD-OLOGICAL
ELEMENTS	Phenomena	Properties	Modes of Treatment
RELATIONS	Patterns	Relations	Comparison Techniques
EMBEDDING SYSTEMS	Ongoing systems [e.g. human-computer systems	Conceptual Systems (e.g., field theory)	Research Strategies (e.g. laboratory experiment)

Figure 1: *Domains and levels of concepts in behavioral and social science research.*

 McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In Readings in Human–Computer Interaction (pp. 152-169). Morgan Kaufmann.

Summary



TYPE 1 - DEMONSTRATION

- INDIVIDUAL INSTANCES
  - NOVEL EXAMPLES
  - REPLICATED EXAMPLES
- COLLECTIONS
  - CASE STUDIES
  - DESIGN SPACES
- GOING BEYOND DESCRIPTIONS
  - HOW TO SCENARIOS

TYPE 2 - USAGE

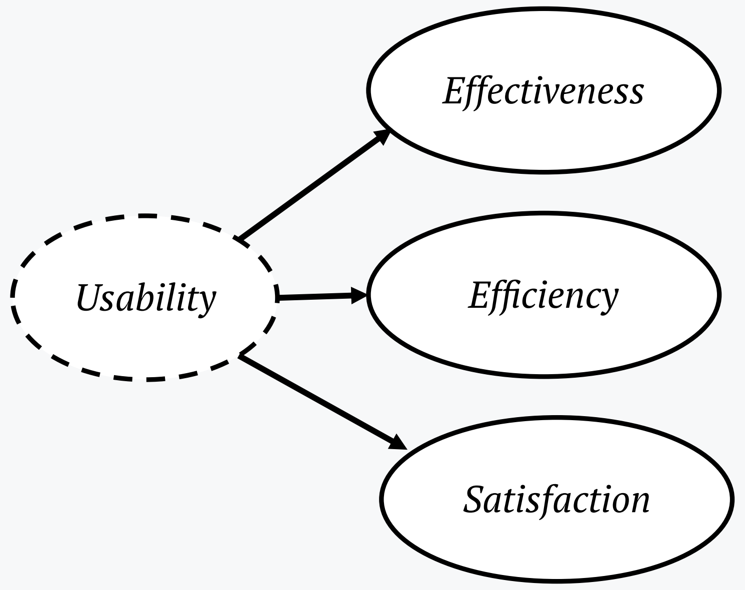
- WAYS TO CONDUCT USAGE STUDIES
  - USABILITY STUDIES
  - A/B COMPARISON
  - WALKTHROUGH
  - OBSERVATION
  - TAKE-HOME STUDIES
- ELICITING USER FEEDBACK
  - LIKERT SCALES
  - INTERVIEWS

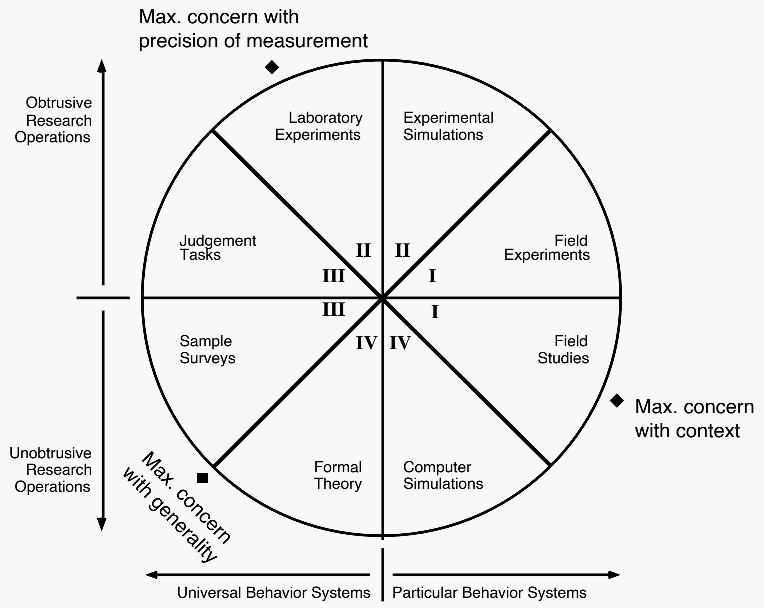
TYPE 3 - PERFORMANCE

- BENCHMARK THRESHOLD
- BENCHMARK COMPARISON

TYPE 4 - HEURISTICS

- CHECKLISTS
- DISCUSSIONS
- TARGETING





Take 5 minutes to answer debrief questions:

This link is also on the talk webpage:
chatw.ch/research-thinking

Limited-time bonus on **Miro**