Methodology

CS 347

Maneesh Agrawala

Announcements

Quiz #5 is worth 2x will take place June 5th (next Wed). **Comprehensive**

No section in week 10 (but discussant commentaries/metacommentaries still due)

AMA for part of lecture on Monday: https://forms.gle/Bb7Y5krrpzmrSQUi7

Last week Mon

Artifacts have politics: the systems we create influence groups and societies, often with undesirable outcomes

Designers hold power: the power is in the theories, methods for ideation, prototyping, and evaluation, and in how we dissipate knowledge.

HCI's role: Identify the ways in which technology shapes the society and envision alternative methods and processes, technical approaches, policies, and designs that mitigate these issues

Last week Wed

Until recently researchers have thought — the space of possible **human behavior has been too vast and complex** to simulate

LLMs show promise in constructing such behavioral simulations

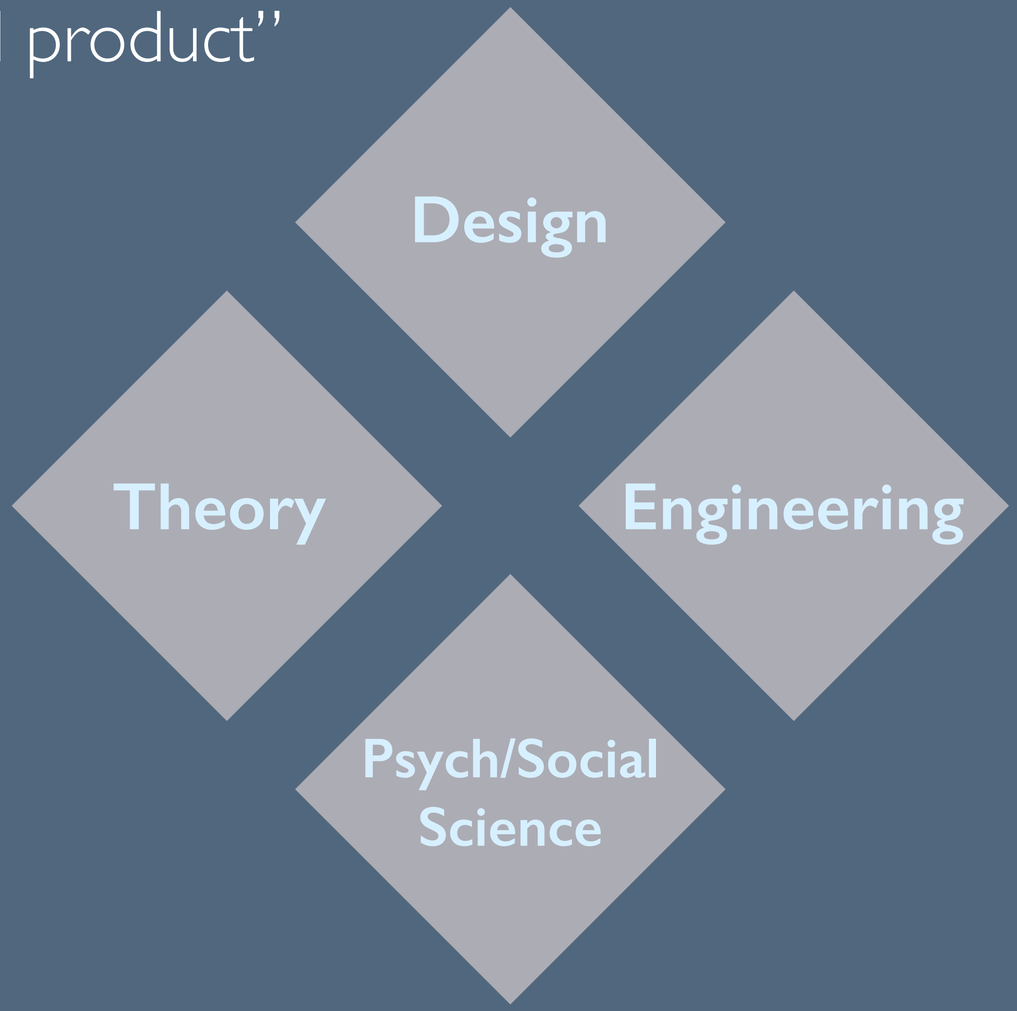
- Emergent behaviors/interactions between agents in **Smallville**
- **Social simulacra** uses an LLM to post in a human-like way to social media to prototype system design

Recall: HCI interdisciplinarity

Before this class: "HCI is design process-iterated product"

After this class:

- An algorithm paper can be HCI
- A design paper can be HCI
- A qualitative paper can be HCI
- A critical theory paper can be HCI
- An EE/ME paper can be HCI
- A field experiment can be HCI
- …



What binds together HCI?

We sometimes think of fields as being bound together by method

Math: formal proof

Applied physics: measurement

Psychology: experiment

Anthropology: ethnography

What binds together HCI?

HCI is not a field that is bound together by method; **HCI is bound together by a shared interest in a topic**

Pro: multiple methodologies present us with many lenses from which we can make progress

Con: it's not always straightforward to know which method to apply

Today

Which is the best method?

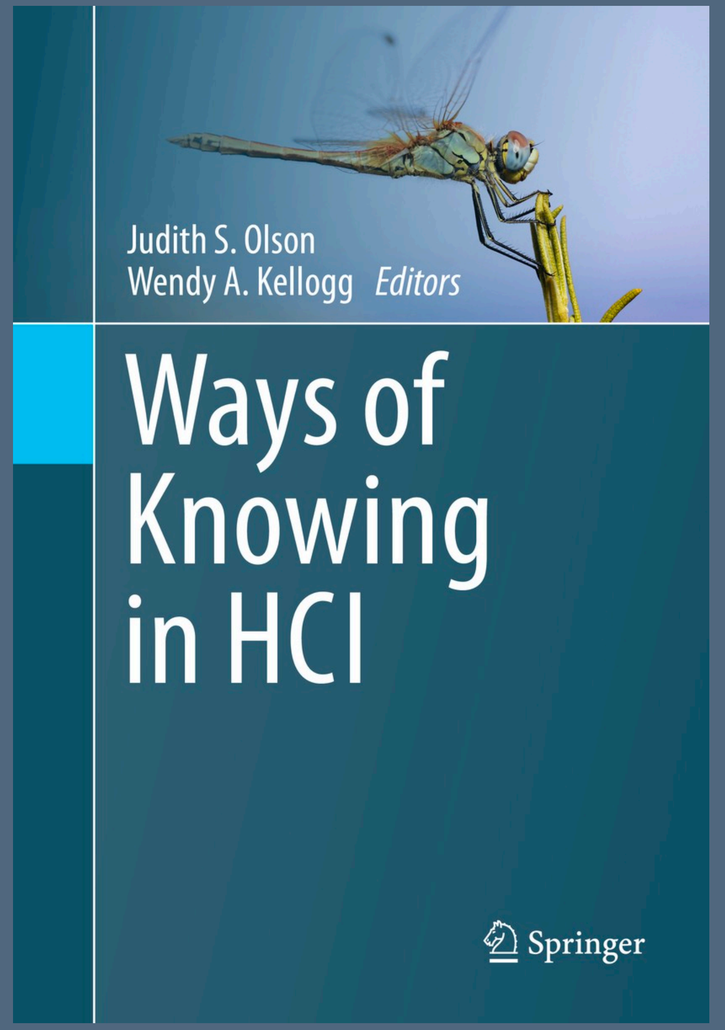
Today

~~Which is the best method?~~

Common methodologies in HCI

How to select your method

Ways of Knowing in HCI



Systems

Systems

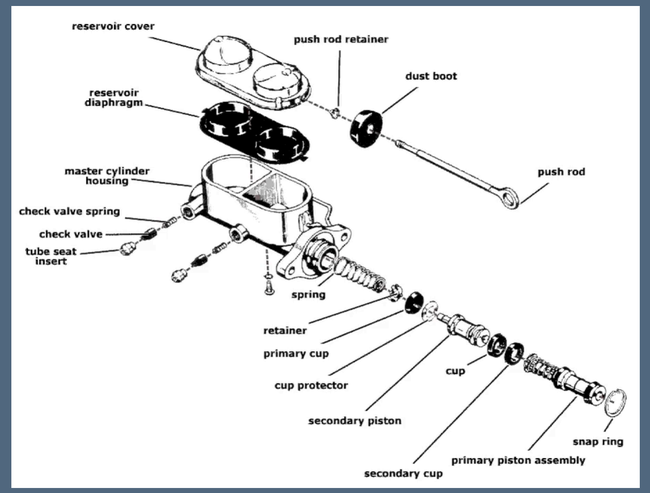
Goal: **develop a novel interactive system that expands the frontiers of what we can create**

Examples from earlier:

Systems

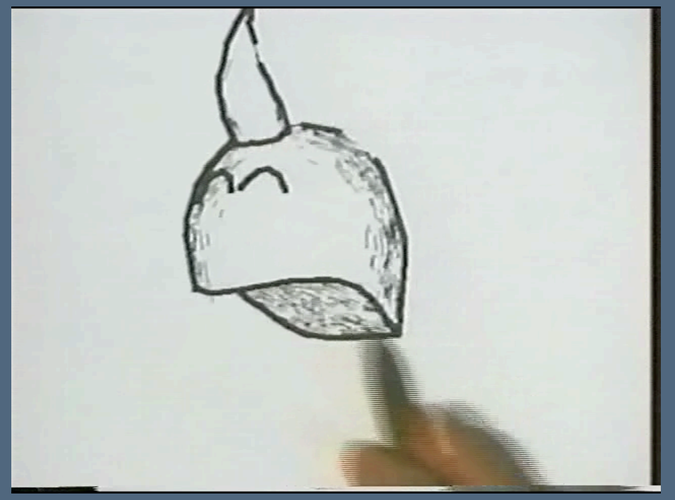
Goal: **develop a novel interactive system that expands our frontiers of how interaction might look**

Examples from earlier:



Interactive Exploded Views

Soli



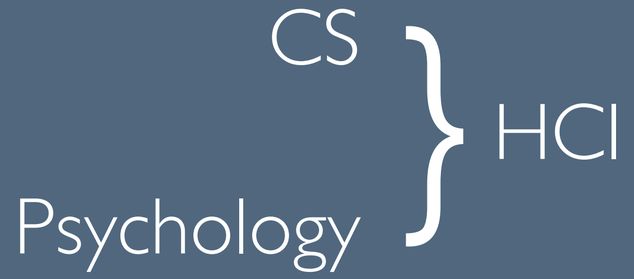
Teddy

Systems

Strength: can inspire and invent new visions of interaction

Challenge: the line between a genuinely new idea and advanced development can become blurry if we're not careful; rarely provides novel behavioral insight

Experiments


CS
Psychology
HCI

Experiments

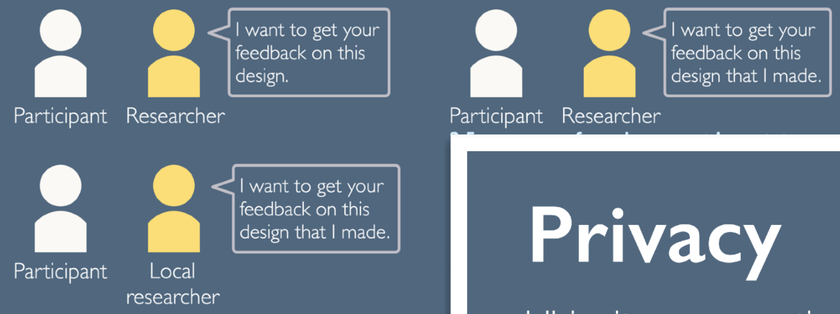
Goal: **demonstrate a causal relationship underlying behavior**

Examples:

Demand characteristics

[Dell et al. 2012]

Response bias due to signals in a study that indicate what the researcher is hoping to see: activating status differences



How does social media impact…

Exposure to diverse political news?

> "We find strong evidence that **[social media] foster more varied online news diets**. The results call into question fears about the vanishing potential for incidental news exposure in digital media environments." [Scharkow et al. 2020]

> "We […] quantified the extent to which individuals encounter comparatively more or less diverse content while interacting via Facebook's algorithmically ranked and further studied users' choices to click through to less cordant content. **Compared with algorithmic ranking, oices played a stronger role in limiting exposure to ontent.**" [Bakshy, Messing, and Adamic 2015]

40

Privacy

Ubiquitous computing naturally raises many questions of how much privacy we are giving up in exchange for its benefits

Behavioral work has documented an empirical **privacy paradox** in which people profess to care strongly about privacy but then willingly give it up in their technology use in practice [Acquisti 2015]

Experiments

Strength: carefully teases out a causal relationship, what affects what?

Challenge: often limited generalizability outside of the experimental context; replicability issues

Correctness issues to be mindful of

- **Internal validity:** is the causal story definitively proven by the method within the frame of the experiment (are there unaccounted confounding factors?)
- **External validity:** do the study results apply to other subjects in a different setting using other measures (are the results generalizable?)
- **Ecological validity:** can you generalize the results to the real world (a subtype of external validity)

Ethnography

Anthropology has joined your party!

Ethnography

Goal: **understand, through participation, how people experience what they do** [Dourish 2014]

Examples:

Industry teams struggle to address these challenges

Ideally, we engage with stakeholders early [Zhu et al. 2018]

But, in practice in industry… [Holstein et al. 2019]

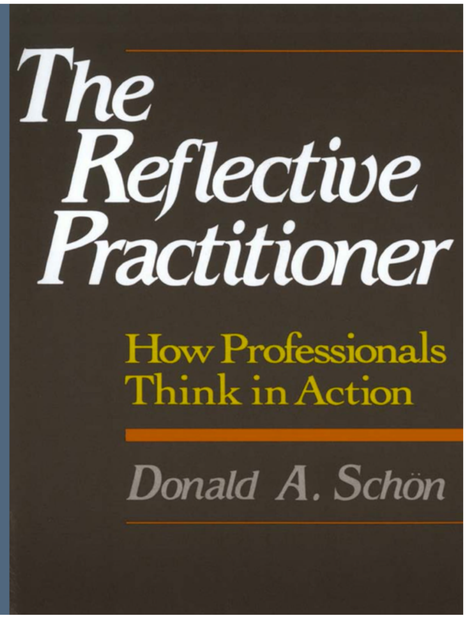
- Data collection is unprincipled ("almost like the wild west") — so if an audit turns up a problem, go collect more training data
- Checklists are difficult, because biases differ by product. Instead, fatalism: "You'll know if there's fairness issues if someone raises hell online."
- Audits require individual-level demographics, but few teams have access to such data

34

Reflective practitioner

How does design work? Why does it work?

Donald Schön [1984] studied a variety of professionals, including designers, and articulated a theory of the how and the why that has remained influential.



Ethnography

Data gathering: participant observation, semi-structured interviews

Analysis: many approaches, but to pick one (e.g. grounded theory)

Strength: "Ethnography revels in particulars" [Dourish 2014]. Aims for generalization rather than abstraction.

Challenge: Not as good a good fit for testing causal theories

Design

CS
Psychology
Anthropology
Design
} HCI

Design has joined your party!

Design

YOU READ THIS

Goal: "The transformation of existing conditions into preferred ones" [Simon 1969]

Integrate behavioral knowledge with technical knowledge to produce a new viewpoint

Examples from earlier:

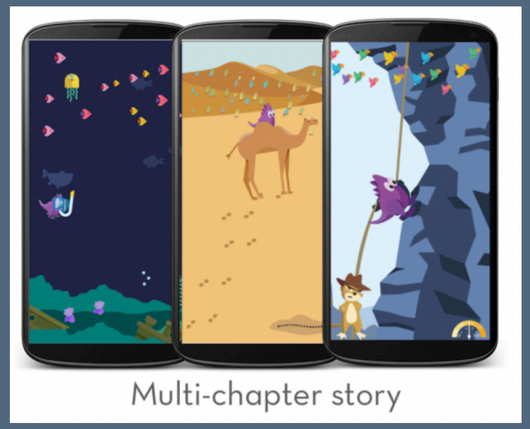
Design

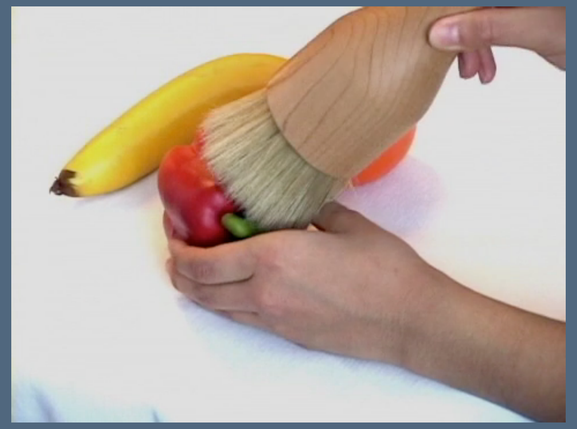
YOU READ THIS

Goal: "The transformation of existing conditions into preferred ones" [Simon 1969]

Integrate behavioral knowledge with technical knowledge to produce a new viewpoint

Examples:





Design

Strength: able to combine diverse elements into a novel whole

> We are still creating something, but now the enabling insight does not need to be technical

Challenge: a combination of elements != a new idea

> To drive a frontier, there must be an a new animating idea or thesis that drives the combination

Computational social science

CS
Psychology
Anthropology
Design
} HCI
Comp. social science

Computational social science has joined your party!

Computational social science

goal: Answer questions about human behavior by drawing on data from social media platforms

(1) A new microscope, can online platforms provide data that enable us to answer longstanding questions in the behavioral sciences?

(2) How has technology-mediated interaction changed our relationships with each other and with the world?

Computational social science

Examples:

How does social media impact…

Our well-being?

> "Receiving **targeted, composed communication from strong ties** was associated with **improvements in well-being** while viewing friends' wide-audience broadcasts and receiving one-click feedback were not." [Burke and Kraut 2016]

Our job hunts?

> "Most people are helped through one of their numerous weak ties but **a single stronger tie is significantly more valuable at the margin**" [Gee, Jones and Burke 2017]

This lecture could have been an email [Cao et al. 2021]

Microsoft researchers investigated their own employees' own multitasking during remote meetings: e.g., are they using Outlook while in a Microsoft Teams meeting?

> Consistently ~30% of meetings involve email multitasking. The odds go up by 2x if the meeting is at least ten people and by 3x if the meeting is ~1hr long
>
> Multitasking does not mean disengagement: often, it's communication with work: "It needs to happen or you can't get all

35

Algorithm audits make these problems visible

Algorithm audit: systematically querying an algorithm and observing its outputs to draw inferences about its opaque inner workings [Sandvig et al. 2014; Metaxa et al. 2021]

Computational social science

Strength: observation and experimentation at scale allow us to execute behavioral research that had been heretofore impractical

Challenge: "Drive-by social science" — analyses that are disconnected from the expertise or theory of the domain experts

(Critical) Theory

CS
Psychology
Anthropology
Design
} HCI
Comp. social science
Humanities

The humanities have joined your party!

Critical theory

Arguments dissecting, probing, and building out the assumptions underlying HCI

Examples:

Feminist HCI [Bardzell 2010]

YOU READ THIS

On one level, feminist theory prompts us to examine how we may be making assumptions about gender or gender roles

> **We ought to view supposedly-genderless constructs (e.g., "the user") as implicitly gendered**
>
> f(ubiquitous computing) → what are we assuming about what sensors people would be willing to wear, or about what kind of sensing and tracking is desirable, that may not apply to non-males?

22

Yesterday's tomorrows

[Bell and Dourish 2007]

Ubiquitous computing is driven not by a technological goal, but by a shared vision of the future.

However, this vision is a future in 1991.

What should the future of ubicomp be, from today's perspective?

54

Critical theory

Strength: can reframe a complex literature into a clearer light, or identify underlying issues that need to be addressed

Challenge: effective synthesis typically requires a broad and deep knowledge of the literature

Tensions in interdisciplinary work

Your stuff is terrible

These methods and fields capture different points of view on how we know things to be true. These can put perspectives in tension:

- We can't trust it if it's not observed in the wild
- We can't trust it if we cannot perform causal inference with a clear mechanism
- We can't trust it if it wasn't measured quantitatively
- We can't trust it if it's not deeply exposed to lived experiences

Rather than taking potshots at other methods, **match the method to the question** - each is best at answering only some questions

Your stuff is terrible.

As CS 347 graduates, you should be able to discard sophistic claims that one of these methods is "good" or "bad" or "always required"

For example, technical and design projects kept getting hammered for poor evaluation. This sucks, since they weren't trying to investigate human behavior

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 academic prototype, it may incorrectly suggest a design's scientific worthiness rather than offer a meaningful critique of how it would be adopted and used in everyday practice. If done without regard to how cultures adopt technology

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

The purpose behind usability evaluation, regardless
actual method, can vary considerably in different co
Within product groups, practitioners typically ev
products under development for 'usability bugs',

Do things precede theory?

[Carroll and Kellogg 1989; Zimmerman and Forlizzi 2014]

Are advances in HCI theory limited by advances in HCI technology?

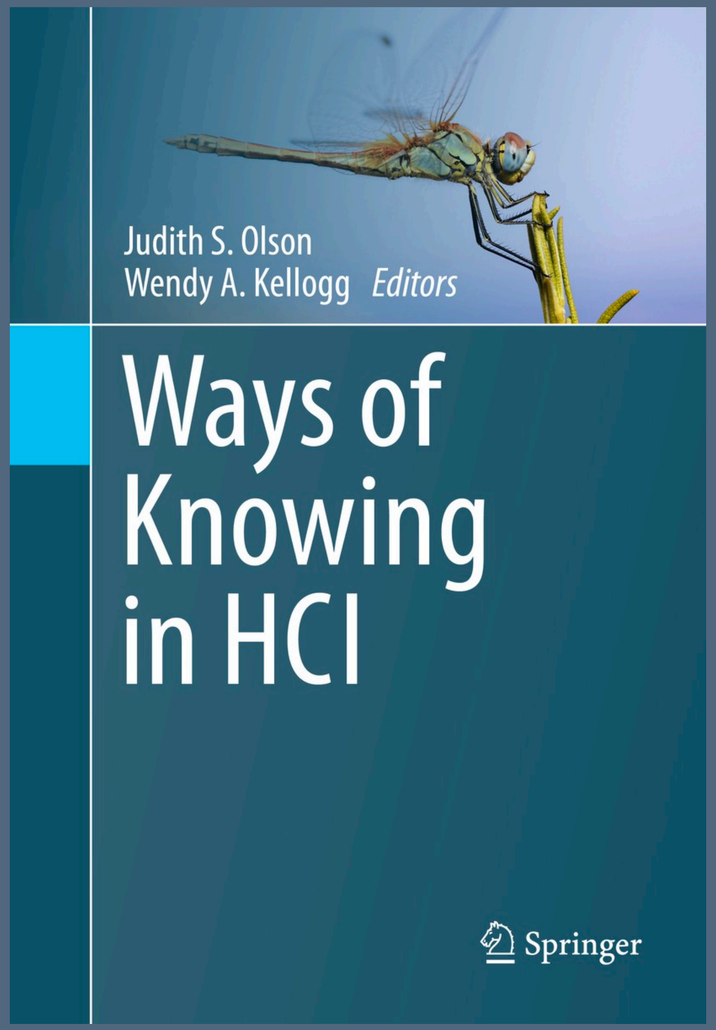
- Sutherland's Sketchpad long predated the theory of direct manipulation
- Engelbart's mouse had to be invented before there could be experimental studies demonstrating that it was a good design
- Each new social media platform launches a raft of new papers

Or are advances in HCI technology limited by advances in theory?

- We had to learn about perceptual psychology before we could explain succinctly how to create effective visualizations

Is our fundamental orientation toward creating new opportunities? Or toward problematizing them? How do we walk both paths (humbly)?

For more…



(Free online while you're at Stanford!)

Reading and Interpreting Ethnography

Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method

Knowing by Doing: Action Research as an Approach to HCI

Concepts, Values, and Methods for Technical Human–Computer Interaction Research

Study, Build, Repeat: Using Online Communities as a Research Platform

Field Deployments: Knowing from Using in Context

Science and Design: The Implications of Different Forms of Accountability

Research Through Design in HCI

Experimental Research in HCI

Survey Research in HCI

Crowdsourcing in HCI Research

Sensor Data Streams

Eye Tracking: A Brief Introduction

Understanding User Behavior Through Log Data and Analysis
Looking Back: Retrospective Study Methods for HCI

Agent Based Modeling to Inform the Design of Multiuser Systems

Designing an evaluation

(Mostly focused on technical and design contributions)

Thanks to the faculty instructor of CS 197 for these ideas

Problematic point of view

"But how would we evaluate this?"

Why is this point of view problematic?

> **Implication:** "I believe the idea is right, but I don't believe that we can prove it."
>
> **Implication:** "Evaluation is distinct from the validity of the idea."
>
> Neither implication is correct. **If you can precisely articulate your thesis, then you can design an appropriate evaluation. If you can't precisely articulate your thesis, then you can't design an appropriate evaluation.**

Step 1: articulate your thesis

A much more productive approach is to derive an evaluation design directly from your idea.

What is the main thesis of your work?

In other words, what do you think is new and matters here?

Prior work

Behavior change can be motivated by quantitative data visualizations

Participatory design brings marginalized stakeholders to the table

Debugging should focus on asking "what is the value of this variable?" questions

Your thesis

Behavior change can be motivated by data-driven narratives

Gaps remain: members of marginalized communities can be alienated by participatory design processes

Debugging should focus on asking "why did this happen?" questions

Step 2: map your thesis onto a claim

There are only a small number of claim structures implicit in most HCI theses. Here are some common ones:

x > y: approach x is better than approach y at solving the problem

∃ **x**: it is possible to construct an x that satisfies some criteria, whereas it was not possible before

x, really? our theory and widely held assumptions would lead us to believe x is true, but we show that x isn't necessarily the case

Prior work	Your thesis	Claim
Behavior change can be motivated by quantitative data visualizations	Behavior change can also be motivated by data-driven narratives	∃ x: narrative visualizations can work (could have been an x > y claim if the thesis implied "narratives are better")
Participatory design brings marginalized stakeholders to the table	Gaps remain: members of marginalized communities can be alienated by participatory design processes	x, really?: participatory design does not live up to its stated potential
Debugging should focus on asking "what is the value of this variable?" questions	Debugging should instead focus on asking "why did this happen?" questions	x > y: debugging through why questions is better than debugging through what questions

Step 3: claims imply an evaluation design

Each claim structure implies an evaluation design

x > y: given a representative task or set of tasks, test whether x in fact outperforms y at the problem

∃ **x**: demonstrate that your approach achieves x

x, really? demonstrate bounds inside or outside of which approach x fails

Your thesis	Claim	Implied evaluation
Behavior change can also be motivated by data-driven narratives	∃ x: narrative visualizations can work	Demonstrate that narrative-driven behavior change has impact
Gaps remain: members of marginalized communities can be alienated by participatory design processes	x, really?: participatory design does not live up to its stated potential	Demonstrate conditions under which PD alienates its stakeholders
Debugging should instead focus on asking "why did this happen?" questions	x > y: debugging through why questions is better than debugging through what questions	Compare debugging through "why" vs. "what" in terms of number of bugs fixed, time, etc.

Summary

HCI's interdisciplinary makes available many methodological orientations. Which to apply depends on your goal. To wit:

Systems: engineer a thing

Experiments: prove a causal thing

Ethnography: understand a thing

Design: craft a thing

Computational social science: analyze a thing

(Critical) theory: think a thing

Design your evaluation by starting back at your thesis, mapping that thesis onto a claim, then deriving the evaluation from that claim

References

Carroll, John M., and Wendy A. Kellogg. "Artifact as theory-nexus: Hermeneutics meets theory-based design." Proceedings of the SIGCHI conference on Human factors in computing systems. 1989.

Dourish, Paul. "Reading and interpreting ethnography." Ways of Knowing in HCI (2014): 1-23.

Greenberg, Saul, and Bill Buxton. "Usability evaluation considered harmful (some of the time)." Proceedings of the SIGCHI conference on Human factors in computing systems. 2008.

Olson, Judith S., and Wendy A. Kellogg, eds. Ways of Knowing in HCI. New York: Springer, 2014.

Simon, Herbert A. "The science of design: Creating the artificial." In The Sciences of the Artificial. (1988): 67-82.

Zimmerman, John, and Jodi Forlizzi. "Research through design in HCI." Ways of Knowing in HCI (2014): 167-189.