Methodology

CS 347
Michael Bernstein
Announcements

Project milestone due Tuesday

Focus on the core element of your idea

Goal: very basic prototype for systems, pilot results+analysis for studies
Last time

Cognitive models create computational proxies of human behavior, to help us characterize and understand how we will engage with a piece of technology.

Model human processor, GOMS, KLM, LLM

Ubiquitous computing requires an understanding of cognition as well: embodied cognition emphasizes how we think with our bodies, whereas distributed cognition emphasizes how we think with the environment.

When our cognition is overloaded, performance decreases.
Recall: HCI interdisciplinarity

Before today: “HCI is design process-iterated product”

After today

- 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 often 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 HCl

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

Examples from earlier:

The Whyline  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
Experiments

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

Examples from earlier:

### 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

- Participant: I want to get your feedback on this design.
- Researcher: I want to get your feedback on this design that I made.

### 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 diverse content while interacting via Facebook’s algorithmically filtered and further studied users’ choices to click through to redundant content. Compared with algorithmic ranking, choices played a stronger role in limiting exposure to content.” [Bakshy, Messing, and Adamic 2015]

### 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
- External validity
- Ecological validity
Ethnography
Anthropology has joined your party!
Ethnography

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

Examples from earlier:

**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

**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 different approaches, but to pick one: grounded theory


Challenge: Not as good a good fit for testing causal theories
Design
Design has joined your party!
Design

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

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

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 $\neq$ a new idea

To drive a frontier, there must be an a new animating idea or thesis that drives the combination (e.g., “ambient narratives’”)
Computational social science
Data science has joined your party!
Computational social science

Draw on data from social computing platforms to answer questions about human behavior

(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 from earlier:

**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 it done.”

**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
The humanities have joined your party!
Critical theory

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

Examples from earlier:

**Feminist HCI** [Bardzell 2010]
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
- (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?

**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?
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
As CS 347 graduates, you should be able to discard sophisticated claims that one of these methods is “good” or “bad” or “always required”.

For example, technical and design projects kept getting hammered around evaluation. This sucked, since they weren’t trying to do behavioral science.
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
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?
<table>
<thead>
<tr>
<th>Prior work</th>
<th>Your thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior change can be motivated by quantitative data visualizations</td>
<td>Behavior change can be motivated by data-driven narratives</td>
</tr>
<tr>
<td>Participatory design brings marginalized stakeholders to the table</td>
<td>Gaps remain: members of marginalized communities can be alienated by participatory design processes</td>
</tr>
<tr>
<td>Debugging should focus on asking “what is the value of this variable?” questions</td>
<td>Debugging should focus on asking “why did this happen?” questions</td>
</tr>
</tbody>
</table>
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
- $\exists x$: 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
<table>
<thead>
<tr>
<th>Prior work</th>
<th>Your thesis</th>
<th>Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior change can be motivated by quantitative data visualizations</td>
<td>Behavior change can also be motivated by data-driven narratives</td>
<td>$\exists : x$: narrative visualizations can work&lt;br&gt;(could have been an $x &gt; y$ claim if the thesis implied “narratives are better”)</td>
</tr>
<tr>
<td>Participatory design brings marginalized stakeholders to the table</td>
<td>Gaps remain: members of marginalized communities can be alienated by participatory design processes</td>
<td>$x$, really?: participatory design does not live up to its stated potential</td>
</tr>
<tr>
<td>Debugging should focus on asking “what is the value of this variable?” questions</td>
<td>Debugging should instead focus on asking “why did this happen?” questions</td>
<td>$x &gt; y$: debugging through why questions is better than debugging through what questions</td>
</tr>
</tbody>
</table>
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

\( \exists x \): demonstrate that your approach achieves \( x \)

\( x, \text{ really?} \) demonstrate bounds inside or outside of which approach \( x \) fails
Your thesis

Behavior change can also be motivated by data-driven narratives

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

Debugging should instead focus on asking “why did this happen?” questions

Claim

∃ x: narrative visualizations can work

x, really?: participatory design does not live up to its stated potential

x > y: debugging through why questions is better than debugging through what questions

Implied evaluation

Demonstrate that narrative-driven behavior change has impact

Demonstrate conditions under which PD alienates its stakeholders

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


