Lecture 5&6:
Surveys & Observations

CS798H: Human Computer Interaction

Interviews are…

- Great…
  - To get detailed responses, and some more from follow-up questions
  - When we know little about a topic and want to explore
  - To garner subjective opinion
  - To seek information about past experiences
- BUT, they are…
  - Limited in sample size (and so less generalizable to overall population)s
  - Costly to conduct
  - Susceptible to biases of interviewers and interviewees (people forget, etc.)
  - Primarily qualitative, hard to gain quantitative insights
  - Dependent on how comfortable interviewee can talk about the subject

Methods: Strengths and weaknesses

	Interview	Survey	Observations
Time per participant	High	Low	High
No. of participants at once	One	High	One
Level of details	Rich details	Not rich	Very rich details
Time covered	Present + past	Present + past	Only what is observed in present
Scope for follow-up questions	Yes	No	No
Truthfulness / reliability	Recall errors / bias	Recall errors / bias	What happens is for the researcher to see

These methods have complementary strengths and weaknesses, so we typically mix methods.
(E.g., interviews + survey, observations + interview)

Why mixed methods?

- Gather Additional / Complementary Data
  - Qualitative vs. quantitative (interviews for what, surveys for how many/much)
  - Observations for "what happens", interviews for "why it happens"

- Triangulation
  - Arrive at the same results through multiple methods/sources/ways
  - Minimizes biases, offsets limitation of one method
  - E.g., geometric constructions, trigonometry, coordinate geometry, rotational geometry could all ascertain some property of circles/ polygons
  - E.g., conduct a larger survey to confirm data from small interview study is generalizable to the larger population.

Triangulation: forms

- Methodological triangulation → use >1 method to answer research question
  - Interview with users + emotion analysis in app reviews
- Data triangulation → gather data from multiple sources / replicate
  - Conduct study elsewhere, prior data from someone else's experiment
  - Ask multiple questions, check if responses are all consistent.
- Theory triangulation → see if data agrees with prior theory, if not you might be missing something (or, if you're lucky, theory needs revision!
  - Done often in physics (didn't agree with classical physics, needed relativity!)
- Participant triangulation → Member checking; confirm with participant.

Surveys

- A direct method for empirical data collection

- Participants provide responses to a list of pre-decided questions

- Typically, takes <10min, compared to interviews that typically last 60-90 minutes.

- Less detailed, but can be sent a lot of people

- Great for quantitative results (e.g., X% of population)

- Great for generalizability (get results from way more people than in interviews)

Conducting surveys

- Start with a research question (just like with interviews)
- Design survey questions (Ensure they answer the research question)
  - Usually closed + some open-ended questions.
  - Not too long (else, no one would fill it in), 5 min ideal, 10-15 min OK.
  - Think about optional and mandatory questions
- Identify how the survey will be conducted
  - Digital / Paper, Telephonic/In person /Simply send and expect to fill in.
- Identify sampling strategy and suitable sample size
- Identify recruitment method

Survey questionnaire

- Keep it short! – Too many questions → fatigue/incomplete responses
- When long is necessary
  - Set expectations, break into sections, show progress, save and continue!
  - Break down / reduce, your scope is probably too broad.
  - Avoid too many consecutive questions of the same kind, to avoid fatigue
  - Most important questions top, so valuable even if incomplete
  - Randomize question order, to ensure you get similar number of questions answered across. (Requires questions not dependent on each other!).
- Motivation → Consider rewards: monetary, raffle, coupons, etc.

Characteristics of survey questions

- Use clear, simple, unambiguous language
  - No scope for clarifications, as with interviews.
  - Where needed, provide descriptions, definitions, examples to clarify
  - Highlight when something is tricky (e.g., if it has NOT, make it bold / caps/underline).
  - PILOTs are your friend! (Also tells how long it takes to fill in, useful in recruitment).
- Ensure you don't exclude some respondents
  - Multiple languages, Accessibility (Braille, screen-reader, color blindness, etc.)
- Ensure relevance of questions
  - Ask questions only if relevant (e.g., based on prior responses)
  - At least don't make it mandatory → Else messes with quality of responses
- GUARD against biases

What kinds of biases?

- Response biases (or biases in participant responses)
  - Acquiescence – saying yes, even when they don't totally mean it
  - Central tendency – choose middle most value most of the time
  - Extreme bias – tendency to choose extremes
  - Ordering bias – order of questions / options causes bias (esp. from fatigue, or leading-in)
  - Recall bias – recall most recent, most common, etc.
  - Social desirability bias: don't want to something that might not be socially acceptable
  - Bias due to power: The person administrating the survey can be seen as authority
  - Satisficing: Pick something more or less, though not exactly what you feel.

Survey questions: kinds

- Open ended questions (tell in 1-2 sentences, can be then put into categories / themes)

- Closed-ended questions
  - Single choice
  - Multiple choice (sometimes useful to say "which top 3" instead of "which all")
  - Rank / order choices
  - Sliders for range of values
    - Likert turn range of values into single choice questions
    - E.g., Never Occasionally Sometimes Often Always
    - E.g., Strongly Disagree Disagree Neutral Agree Strongly Agree

Asking closed-ended questions

- What to put in options?
  - Reuse questions from other surveys, with modifications as needed.
  - Seed with results from other methods (e.g., from interviews / prior research)
  - Always add an Other / Maybe / Don't know (Plan for surprises!).
  - When using Likert scales, use standard language
  - Rethink whether you need 5 or 7 (too many options can be confusing)
    - Consider defining / clarifying(e.g., Always = 5+ times a day, Often=1-2 times a day, Frequently = a few times a week, etc.)
  - Rethink whether you need odd / even (central tendency bias)

Guarding against biases

- Randomize options within a question, randomize question orders
- Provided an option for Other/None of the above / All of the above
- Ask questions on both ends of the spectrum, to make either side socially acceptable
- Anonymity
- Avoid leading in:
  - E.g., "Social media is bad for older adults." – Agree or Not
  - Replace with a pair of questions:
    "How much do you think social media is helpful for older adults" +
    "How much do you think social media is harmful for older adults"
  - You could then use: A great deal, Much, Somewhat, Little, Never for each of those.
- Hard question – middle value or not in likert scales.

Sampling

- You identify the population of interest, and sample from them.
  - Include everyone → Census
  - Probabilistic sampling → Known probability of individual being included in the sample
    - Simple random → Randomly invite members to participate (ensure they satisfy inclusion/exclusion criteria)
    - Systematic → Every kth member
    - Stratified → Break into subpopulations, and randomly sample from them
    - Cluster → Break into subpopulation (where each cluster looks like the population itself!), and then draw from them.
  - Non-probabilistic
    - Convenience
    - Voluntary responses
    - Purposive → Decide who, for good reasons
    - Quota → X% from each subpopulation (usually representative of entire population)
    - Snowballing → not so much in surveys, as with other methods

Sample size

- Until saturation
  - After some responses, the overall data we get saturates & stabilizes
  - Ensure diversity in demographics (sanity check!) & then stop!

- External constraints (time, budget, deadlines etc.)

- Probabilistic: depending on margin of error & confidence level
  - Written as z % ± e (Out of 100 repetitions of survey, error within e, Z% of the time)

  - Sample Size = $\frac{K}{1+KN}$, where K= $z^2 \cdot \frac{P(1-P)}{e^2}$
  - P is a probability of a certain response from population, typically set to P=0.5

- How many to recruit: Sample size / expected response rate * 100

Recruitment channels

- Emails, social media, SMS, posters → self-selection bias

- What about people without tech / not registered?

- Pick people from telephone directory → what about those without phones?

- Consider recruitment / sampling on events, people, organizations as well!

Observations

Observation Methods

- Direct method
- Simply observe people doing something
- Advantages:
  - Mitigates recall bias
  - Captures minor subconscious things people might do (e.g., typos!)
- Disadvantages:
  - Hawthorne effect bias
  - Time consuming
  - Only in time, not past history. Similarly, not why (directly).

Conducting observational studies: same drill!

- Research questions
- What to observe?
  - What do you want to observe (e.g., learning to use a new app)
  - What exactly (app you give, or an app they choose and want to)
- When?
  - When they do it anyway (contextual inquiring)
  - When you tell them to (for the study)
- Where?
  - In a lab setting (in-vitro)
  - In the field (in-vivo)
- How long?
  - Entirety of task? fixed time? Can participants abandon task midway?

Drill continued

- Participants
  - Should be sampled from the population of interest

- How many
  - Typically, until saturation, sometimes happens at 5 and a few more.
  - 8-15 is common (provided it saturates)

- What kind of data you get?
  - Qualitative (in the form of video/screen + audio recordings, or notes)
  - Just record anyway!

Minimize biases

- More than one observer
- Observe from outside
- Tell participants it is OK to make mistakes, you are interested in their way, there is no right or wrong way
- Distance yourself from the task/product (tell them you are helping the team understand, not that you are part of the team!)
  - If you can't, hire someone to do it for you!
- Pick tasks from the real world!
  - Confirm if the tasks are reasonable post-study
- Triangulate – with post-study interviews, surveys, etc.

Useful variations

- Contextual inquiry
  - Recruit "next time you do X, can you call me?" and run with your gear and notebook.
  - Note: Also useful to administer specialized surveys or interviews, sometimes
- Think-aloud
  - Ask people to talk along what they are doing (helps understand why might be doing what they are doing)
  - Alternatively, say you are an apprentice, and ask them to teach you.
  - Warning: think-aloud interferes with natural flows of thinking, so use with caution.
- Observe, make notes of times, and go back and ask follow-up questions
  - To add more details when not clear / something interesting happens

Methods: Strengths and weaknesses

	Interview	Survey	Observations
Time per participant	High	Low	High
No. of participants at once	One	High	One
Level of details	Rich details	Not rich	Very rich details
Time covered	Present + past	Present + past	Only what is observed in present
Scope for follow-up questions	Yes	No	No
Truthfulness	Recall bias	Recall bias	What happens is for the researcher to see

Readings

- Lazar, Chapter 7 (Edition 4) on data gathering methods
  - Covers surveys, interviews, observations

- Additional reading:
  - Lazar (link posted on Resources)
  - Skim through chapters on interviews, surveys and observations
  - Great for reading on biases, sampling, advise on designing good questions.
  - Highly recommended for CGS minors, PG students in Design/CGS/CS, or for anyone interested in usability, HCI, entrepreneurship and product management.