Looking Back: Retrospective Study Methods for HCI

Helen Wauck, Yi-Chieh Lee
HCI Data Collection in Short Term Studies

In HCI, data collection is generally in the moment - think-aloud protocols, logs, surveys, interviews

“In the moment” - get participants’ immediate reactions before they forget

Works well for short term studies (60-90 minutes)

But what about longer term studies?
HCI Data Collection in Longer Studies

Typically use specialized “in the moment” data collection methods:

- Diary studies – participants directed before study to record events periodically
- Experience sampling – participants directed by periodic external signal to record events during study
- Log data (see Ways of Knowing chapter)
HCI Data Collection in Longer Studies

However, these approaches have some significant problems:

- Log data analysis cannot provide **context** and **motivation** for user behaviors

- Diaries and experience sampling:
  - Participants lose motivation over time
  - Hawthorne Effect – worker productivity study
  - Observer-expectancy effect – Clever Hans

- So how do we address these limitations?
An alternative approach: Retrospective Methods
(also known as retrospective cued recall)
What are retrospective methods?

1. Record data about participants’ behaviors over time
2. Review data with participants afterward to provide context and motivation

Why would this be useful? A concrete example:

<table>
<thead>
<tr>
<th>Player ID</th>
<th>Timestamp</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0:30</td>
<td>NEW GAME</td>
</tr>
<tr>
<td>1</td>
<td>0:31</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:35</td>
<td>NEW GAME</td>
</tr>
<tr>
<td>1</td>
<td>0:40</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:45</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:55</td>
<td>NEW GAME</td>
</tr>
</tbody>
</table>
Strengths/Weaknesses of Retrospective Methods

Strengths:

- No interruption – more naturalistic behavior
- Less burden on subjects – more motivation to complete study
- Provide context and motivation for behaviors

Weaknesses:

- Memory biases or inaccuracies, false memories
The Key to Effective Retrospective Methods

How do we limit the weaknesses (memory bias, inaccuracy) of retrospective methods?

The key to designing effective retrospective methods: Cues.
How Human Memory Works

Human memory is fragile, biased, and inaccurate. We tend to:

- Reconstruct memories using “prototypical pattern” instead of actual events
- Take cues from the researcher’s questions
- Associate past events with each other, emphasizing similarities

These tendencies often result in false or inaccurate memories.
How Human Memory Works

However, we can increase accuracy of recall using good memory cues:

When given sensory prompts (data, sound, imagery), we are very good at recognizing past situations we’ve been in

Images are particularly powerful cues

Good cues provide participants with the relevant context: time, place, and activity

Choosing good cues (especially visual ones) is key to making retrospective methods work
Types of Retrospective Methods

- Logging
- Video
- Eye tracking
- Day reconstruction
- Image-based experience sampling and reflection
Logging

Comprises one half of retrospective methods – unobtrusive behavior tracking

Recall previous example:

<table>
<thead>
<tr>
<th>Player ID</th>
<th>Timestamp</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0:30</td>
<td>NEW GAME</td>
</tr>
<tr>
<td>1</td>
<td>0:31</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:35</td>
<td>NEW GAME</td>
</tr>
<tr>
<td>1</td>
<td>0:40</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:45</td>
<td>MOVE</td>
</tr>
<tr>
<td>1</td>
<td>0:55</td>
<td>NEW GAME</td>
</tr>
</tbody>
</table>

See the Log Data chapter in Ways of Knowing for more details
Video

Akers et al 2009 studied “critical incidents” in the use of 3D modeling software.

Automatically logged critical incidents (undo and erase events):

https://www.youtube.com/watch?v=k2rULFQGcDg

Showed participants video clips from 20 seconds around each incident

Video clips provided visual and temporal **context** for participants to explain **why** each incident was noteworthy (or was not)
Eye tracking

Used to understand where participants’ attention is focused on an interface

Overlaying eye tracking capture on top of video/screen capture gives participants enough cues to recall their visual motion and the reasons behind it

Can also be used to see if participants noticed a feature at all
Day Reconstruction Method

Participants reconstruct daily experiences as a series of episodes.

Chronological, episodic structure allows participants to draw on episodic memory.

Typically done at the end or beginning of each day (avoids asking participants about their beliefs about the entire, global experience – more on this later).

**Use:** when participants understand the nature of the study.

**Don’t use:** when research questions concern experience across multiple days.
Image-based Experience Sampling & Reflection

A type of experience sampling that also incorporates retrospective methods:

1. Automatic still image capture
2. Post-study participant reflection with still images
Evaluating Retrospective Methods in HCI

1. Data collection:
   a. How will the data be collected and what kinds of data will be collected? Automatically? Or will it be collected by manual intervention? To what extent will manual annotation interfere with the actual behaviors under study?

2. Study duration:
   a. How long will the study run? Longer runs have the advantage of collecting larger amounts of data, and thus have a higher chance of observing events of real interest, but this interacts with longer term biasing effects
**Time Span**

1. **Short-Term studies (<2h):** Think aloud protocol, retrospective is gathered immediately after task.  
   a. Often with the retrospective gathered by a think-aloud protocol as the participant observes a playback of the actual study as captured by **video recording** of the participant, their screen behavior, or their **eye movement behavior**

2. **Intermediate-term studies (> 2h, < 2 days):**  
   a. Studies that run for 1 or 2 days can be naturalistic in ways that the short-term studies are not, because they are conducted in lab settings under tight time constraints

3. **Very long-term studies (>1day):**  
   a. **Daily check-in** protocol as a way to keep in touch with their participants,  
   b. “irresistible tendency for subjects to clean up their act and to describe a more coherent and well-thought-out strategy than is normal”
Evaluating Retrospective Methods in HCI

1. Review instruments:
   a. How will the participant and the researcher review the data? Usually some kind of playback system is needed to select salient episodes or events from the data stream.

2. Sampling frequency:
   a. What data sampling rate should be expected and what events cause the data to be collected? Will it be random time sampling, event driven (e.g., by a user action being taken), or periodic (e.g., every hour or at the end of the day)

3. Delay of review:
   a. When will the data be reviewed with the participant? Periodic reviews are useful for longer term experiments, but it becomes difficult to avoid giving the participant subtle clues about what kind of behaviors are the “right” ones
A Sample of Retrospective Analysis - IE Capture

- How participants thought about and framed questions as they went through their research process over hours, days, and weeks.
  - By their nature, such tasks are difficult to capture in laboratory settings
  - **IE Capture** was a browser add-on that captured not just moments in a user’s behavior of a Web browser but also, crucially, complete screen images
Figure 5. IE Capture Viewer—a tool for reviewing the participant's log and screen images for discussion and retrospective cueing. The participant's screen image is visible in the center of the display, with the stack of windows present at the time of screen capture, an essential part of cueing for long-term recall. The lists on the right hand side are for quickly moving among the log events and captured images for discussion purposes with the participant.
Methodology Design

– **Data collection**: complete screen captures, URL, time-stamp; triggered by the completion of the loading of a Web page in the browser.

– **Study duration**: varying from 1 to 6 weeks (mostly 2 weeks in length).

– **Review instruments**: the collected screenshots were reviewed with a custom-built data viewer that allowed the participant to browse forward and backward in time through the collection.

– **Sampling frequency**: samples were collected on-event (at document-load time) whenever the participant was using Internet Explorer as their browser.

– **Delay of review**: 1–6 weeks (most often 2 weeks) after data collected.
Methodology Design

During the interview,

The participant would review the collected series of screen captures

Providing the backstory in response to questions asked by the interviewer
Interview Questioning Procedure

1. **What happened next in the search process?**
2. The participant was instructed to describe the next event if they felt “reasonably confident” that they knew what happened,
   a. in particular, focusing on what search terms were used, and whether or not that particular next search was successful.
3. **If the participant could not recall,**
   a. then the experimenter would go forward in time, showing them one event image after another, pushing forward in time, until the participant could recollect what was going on and was able to predict what the next search event would be.
Validity

1. The number of correct predictions based on a cued recall.
2. The number of times they had to go to a previous page before they could recollect what was going on in the search.
   a. A good recollection happens when the participant can accurately recall the next search event after just one or two “cue” screen images.
Fig. 5  Events farther in the past required more and more pages to accurately recall the next search event. After about 6 days out, a participant usually needed around 2 or 3 pages to remember what happened next (The numbers are non-integers as they represent the average number of pages required by all study participants)
Pragmatic Guidelines

1. Choosing good cues: It is important to capture data that will provide useful memory cues
   a. Image and screen capture
   b. Contextual detail

2. Walkthrough methods of interviewing:
   a. present earlier data and maintain the **chronology of events**
   b. Be aware that the cueing stimuli used are the only stimuli being tested for recall. Seeing **additional cues may significantly alter the answers** (and improve!) to later interview questions.

3. Asking for predictions
   a. validate the accuracy of the recollections being elicited by the cues
   b. “ After you saw this screen, what was your next action? “
Pragmatic Guidelines

1. Face-to-face interviews:
   a. face-to-face connection between the participant and the researcher is more effective than distance methods
   b. Social desirability response biases are more likely to take place in telephone interviews than in face-to-face interviews

2. Ways to avoid the false memories effect:
   a. Interviewers can easily (and often accidentally) introduce false memories by the way they frame their questions.
   b. Avoid direction about what parts of the behavior should be noticed,
   c. Avoid value statements about the behaviors in question
   d. Avoid asking for global affective responses from experiences in the past

3. Avoiding testing children:
   a. Cognitive limits on younger children’s ability to reflect on previous performances
Lessons learned: My Personal Experience

Fall 2015 study conducted during CS465: User Interface Design

Research Question: What are the tradeoffs of soliciting design feedback from peers in the class versus different online crowds?

- **Course week 9**
  1. Upload early stage prototype
  2. Get external feedback
  3. Give peer feedback

- **Course week 10**
  1. Rate feedback quality
  2. Revise early stage prototype

- **Course week 11**
  1. Submit revision
  2. Select actions
  3. Complete survey
Post-study interviews

Occurred about 1-2 weeks after the last day of class

Asked participants to reflect on their experience getting and giving feedback:

- What do you think is the greatest benefit of getting feedback from external crowds? From peers?
- Which source would you prefer to get feedback from in future courses?
The Problem: Lack of Cues

Participants were not shown the feedback they received during the interview.

Participants were not shown which source feedback came from during the interview.

So, participants interpreted questions in context of the feedback they got, BUT:

- Had difficulty recalling anything specific about feedback
- Too confident about feedback’s source when they did remember
How I could have fixed it

During interview, change procedure to:

1. Show participants the feedback they received
2. Ask them which feedback they thought came from each source
3. Ask them which source they thought was the best
4. Reveal source of each piece of feedback
5. Ask for reactions/comparison to expectations
6. See if reveal changed their answer to #3