When Participants Do the Capturing: The Role of Media in Diary Studies

Scott Carter
EECS Department
University of California, Berkeley
Berkeley, CA 94720
sacarter@cs.berkeley.edu

Jennifer Mankoff
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
jmankoff@acm.org

ABSTRACT
In this paper, we investigate how the choice of media for capture and access affects the diary study method. The diary study is a method of understanding participant behavior and intent *in situ* that minimizes the effects of observers on participants. We first situate diary studies within a framework of field studies and review related literature. We then report on three diary studies we conducted that involve photographs, audio recordings, location information and tangible artifacts. We then analyze our findings, specifically addressing the following questions: How do context information and episodic memory prompts captured by participants vary with media? In what way do different media "jog" memory? How do different media affect the diary study process? These questions are particularly important for diary studies because they can be especially useful as compared to other methods when a participant *intends* to do an action but does not or when actions are particularly difficult to sense. We also built and tested a tool based on participant and researcher frustrations with the method. Our contribution includes suggested modifications to traditional diary techniques that enable annotation and review of captured media; a new variation on the diary study appropriate for researchers using digital capture media; and a lightweight tool to support it, motivated by past work and findings from our studies.

Author Keywords
Diary studies, qualitative methods

ACM Classification Keywords
H5.m. Information interfaces and presentation (*e.g.*, HCI): Miscellaneous.

INTRODUCTION
Researchers have a handful of tools and techniques available for understanding everyday human behavior. But many of these techniques either require significant time and resource investment by researchers, such as contextual inquiries, or are divorced from empirical evidence, such as surveys. The diary study is a method of understanding participant behavior and intent that attempts to manage this gap by having participants record events as they happen. This recording usually occurs in one of two ways: participants answer predefined questions about events (feedback studies) or participants capture media that are then used as prompts for discussion in interviews (elicitation studies).

Field studies that require the researchers' persistent presence are difficult to scale. On the other hand, because of their reliance on participants to collect data, feedback studies have the potential to be scalable. However, participants are often reluctant to use them because the act of answering questions is a significant distraction from their main task. Also, because of the lack of an objective observer there is no way to verify to what extent logged information matches actual events. Media elicitation studies mitigate both of these concerns. In a media elicitation study, participants capture events, usually by taking a photo, and are asked about the event during an interview at a significantly later point in time. Thus for elicitation studies, capture is quick, and while the captured media still represents a subjective point-of-view, it has some empirical value. Barsalou posited that episodic memory can be improved when a person is presented with cues about an event such as who was involved, where it occurred or what was done just before and after the event [3]. However, while researchers have recently begun using diary studies using photo-elicitation, it is not evident how well media capture these cues and to what extent media facilitate participant reconstruction of events. Also, different media *types* will likely evoke different reconstructions and attitudes towards an event, but no study has yet shown how.

Based on these concerns, one contribution of this paper is a set of suggested improvements to the diary study technique, derived from three studies of the technique itself in action. For two of these studies we played the role of a participant observer by involving ourselves in an ongoing study. Specifically, we observed the process of using the method, analyzed results from the study and interviewed the researchers involved about their experiences. The other study we ran ourselves to gain first-hand insight into the issues involved in running a diary study and to compare and contrast the use of different capture media: photos, audio clips and tangible (physical) objects. While photo diary studies are gaining in popularity, use of the other two media is limited.

Our studies revealed a need for situated annotation of captured events in elicitation studies. We found that the best approach to feedback studies may be to combine media capture

Scott Carter, Jennifer Mankoff: When participants do the capturing: the role of media in diary studies. CHI 2005: 899-908

Figure 1. A selection of tangible objects collected by participants in the festival study. "The flowers (upper left) ... mirror how I think about jazz."

with structured, question-and-answer based annotations. Our studies also revealed the usefulness of different media in different situations. Specifically, we found that images lead to more specific recall than any other medium, but that audio, in addition to making it easier for participants to capture information that does not have a visual representation, can be used clandestinely in situations in which participants do not feel comfortable using a photo to capture an event. We found that information about location does not significantly impact recall, and that tangible objects are more likely than other media to prompt discussion of broad attitudes and beliefs (Figure 1). We also noticed unforeseen issues in elicitation interviews. For example, while media capture lent itself to a sequential review of data, interview discussion tended to follow themes, causing problems for participants and researchers when they referenced captured data out-of-sequence.

Another contribution of this paper is a technique and a tool to support media-based diary studies. Our experience with media-based diary studies as well as reports in the literature, indicate that it is important to mitigate the impact of a study on participant's everyday interactions and encourage participant recall of ambiguous data. We also found it important to provide support for interview preparation. To address these issues we propose a diary study pipeline that borrows from both feedback and elicitation methods to maximize participant recall and interview preparation while minimizing situated logging. We then built and tested a lightweight tool, *Reporter*, to support this pipeline. Results showed that participants were able to learn the tool rapidly.

THE DIARY STUDY METHOD

The diary study is a method of understanding participant behavior and intent *in situ* that minimizes the effects of observers on participants. Diary studies differ from other field study methods in that researchers are remote from participants and participants control the timing and means of capture. When researchers are local with respect to participants, as in contextual inquiries, they are able to discuss the implications of events and actions with participants immediately. These studies yield data less pigeonholed by a participant's particular perception of an event but are subject to presentation effects (*i.e.*, participants may act differently because of the presence of the researcher) and are time-consuming and difficult to scale. Also, when researchers control capture they are able to obtain objective data about participant's activities but do not necessarily gain an understanding of the events that are important to the participants. An example of such a method is Experience-Sampling Method (ESM), in which participants are interrupted throughout the day to answer a set of questions.

Diary studies can be broken down into those that use media captured by participants as prompts for discussion in interviews (elicitation studies) and those that require participants to answer predefined questions about events (feedback studies). Feedback studies may also require participants to capture media to serve as prompts, but the principle difference between elicitation and feedback studies is that elicitation studies involve *synchronous* communication between researcher and participant (*e.g.*, interviews) while feedback studies involve *asynchronous* communication between researcher and participant (*e.g.*, questionnaires). In some studies the methods are combined, with results from feedback serving as prompts for discussion during the elicitation study.

Another difference between feedback studies and elicitation studies is that in feedback studies participants should provide information about an event immediately after they perceive it, whereas in elicitation studies participants only capture some aspect of an event when it occurs and provide information about it later during interviews. Thus, a typical feedback study will ask participants to answer questions about some event as soon as it occurs, whereas in an elicitation study participants merely capture some information about the event that will serve as a memory cue during a later interview. Feedback studies have the drawback of potentially overburdening participants with questions, especially when the number of events reported is high [21]. Because participants can rapidly capture prompts, such as a photograph, elicitation studies tend to be much less burdensome. But because questions are asked at the time of the event, or *in situ*, feedback studies are more likely to provide accurate responses to questions that depend on recall of the event. Thus, the two methods represent a tradeoff made between accurate recall but burdensome logging (feedback) versus potentially inaccurate recall but unobtrusive logging (elicitation).

RELATED WORK

Social science researchers and technologists have recently begun running diary studies that involve participants capturing media (*e.g.*, [4, 6, 18, 19, 23, 24]). Researchers have also begun to develop techniques to aid the use of technology in related methods, such as ESM [2, 15]. However, in participant-driven diary studies the impact on participant responses of using different media has not been evaluated.

Participant-driven diary studies

In this section we review participant-driven diary studies, or those in which participants, rather than researchers, use media to capture events. Feedback studies are inherently participant-driven: participants must respond to questions using some medium when an event occurs. However, researchers in both behavioral and technological research communities have only recently begun to explore participant-driven elicitation studies.

Participant-driven feedback studies usually rely on paper-based forms as the feedback medium [7, 11, 22, 25]. However, one of the concerns with using paper-based feedback studies is sustained subject participation. In an attempt to address this issue, Palen and Salzman experimented with cell phones as a feedback medium [19]. From their studies they derived ways to encourage subject participation, including the use of periodic reminders and reimbursement strategies. They also recommend that researchers provide participants with feedback about the level of detail of their responses.

Participant-driven elicitation studies are rare with the exception of photo-elicitation studies. Brown *et al.* used photo-elicitation to understand design requirements for information capture devices, and O'Hara *et al.* used the same method to understand transaction decisions [4, 18]. In his study of young Buddhist monks in Sri Lanka, Samuels compared participant-driven photo-elicitation to word-only interviews and found that participants were far more detailed in their description of everyday events with photo elicitation [24]. According to participants, the difference arose because the photos that they took had more "meaning and value" to them and that they could "explain more when the pictures are close at hand." Samuels also found with photo-elicitation that participants were better able to make novel associations among tasks and that participants tended to remain more focused on the interview.

Clark-Ibáñez also used participant-driven photo-elicitation in her studies of children attending elementary schools in urban environments [6]. She described that background information in photos can often be crucially important, specifically citing a case in which participants discussed the "'tagging' of gang names and symbols" in the periphery of one image that she did not initially notice. She also found that photos tended to hold participants' attention and found them useful for structuring interviews. However, as she developed the photos herself, participants could not review and potentially remove photos before the elicitation interview, frustrating some.

Sampson-Cordle used participant-driven photo-elicitation to construct photo essays, or combinations of interview transcripts and photos about a related topic, in her study of a small rural school and the community in which it was situated [23]. She found that participants would often take pictures of similar objects but have vastly different reasons for taking the photo. She also found it vital to allow participants to erase photos.

In summary, feedback studies using a medium more convenient for participants, such as cell phones, may yield higher use rates. Also, recent use of photo-elicitation has shown it to be a promising method of gaining more detail about participant's everyday events, augmenting participant focus on the interview itself and encouraging participants to make new associations. This work has also shown the importance of peripheral information in photographs as well as identifying the need for participant review of photos prior to conducting the interview.

Unobtrusive techniques for field studies

In this section we compare and contrast the diary study method to two other often used non-intrusive techniques for understanding participant behavior in everyday scenarios: cultural probes (and the related method technology probes) and ESM.

In the cultural probe method researchers design a set of tools that participants use to express their feelings, beliefs, and attitudes. Considerable planning goes into the design of the probes themselves to make use of familiar functionality (such as a camera) while encouraging participants to examine their daily life in a way that have not before (such as asking them to take a photo "at 8pm on a Sunday") [12]. Cultural probes are similar to diary studies in that they are highly participant controlled, but are intended to capture general attitudes and social trends rather than everyday interactions. In a related technique, Hutchinson *et al.* introduced technology probes, or simple tools designed to encourage creative use, to a field situation to generate design requirements for more specific tools [13]. However, technology probes are not designed to capture everyday interactions.

Using ESM, participants are interrupted throughout the day to answer a set of questions. Interruptions may occur when a predetermined amount of time has passed or when an activity or event occurs [2, 15]. While ESM is useful for measuring the amount of time participants spend on everyday tasks, it is not as useful when one is interested in events in which a participant was unable to do something because of a limitation, or when a participant action is difficult to sense. In these cases the diary study method is more appropriate because it allows the participant to control when to answer questions or log an event. Also, unlike ESM but similar to cultural probes, data from diary studies may implicitly indicate the importance a participant attributes to artifacts, people, and places.

STUDIES

None of the studies cited above concentrate on how using different media for communication and prompting might affect the diary study method as a whole, including the types of responses different media elicit as well as how different media effect the process from a researcher's perspective. To thoroughly explore media use in diary studies we analyzed the results of three studies to better understand how to support media elicitation and diary studies. The studies included an elicitation study in which participants used photos as prompts; a hybrid phone feedback and elicitation study; and an elicitation study comparing photos, audio and tangible objects as prompts. For the first two of these studies we

played the role of a participant observer, observing and analyzing studies run by other researchers. We ran the third study ourselves.

Photo diary in an everyday setting

This study was run by another group at our institution, the members of which we refer to as *researchers* below. We were involved in the study from beginning to end, observed the process of using the method, analyzed results from the study and interviewed the researchers involved about their experience. In the study, researchers explored how people search for, consume and produce information. The researchers' *participants* in the study captured information consumption or production events using digital cameras and used photo-elicitation to explore the meaning of captured photos.

Method

After piloting the study with eight participants, the researchers recruited an additional 11 participants for this study from within the social networks of the researchers. The participant group was split evenly between men and women. Ninety percent of participants reported some college education and sixty percent some graduate education.

The researchers gave each participant some basic instructions and a digital camera. The digital camera (Logitech's 961305-0403 Pocket Digital 130) was chosen as it is small, robust and easy-to-use. The researchers asked participants to take a photo each time they noticed that they were consuming or producing information. An instruction sheet noted that things like reading a newspaper, surfing the web, reading email, watching television, listening to the radio and other similar activities were of interest and should be photographed. Also, the researchers asked pilot participants to create written annotations for each photo on a small notepad. However, as all of the pilot participants abandoned written annotation after only a few attempts, the researchers decided not to ask for such data of the study participants.

Participants captured photos for an entire day from the point they awoke until they retired in the evening. Because of the amount of time required to interview participants about rich photographs the researchers felt it important to discourage casual picture-taking, and thus instructed participants to think of the device not as a camera but rather as an information-event capturing device. One day after the day chosen to diary, the participant was interviewed for approximately one hour. The participant first completed a small demographic survey that included questions about their information use. The interview was qualitative in nature and revolved around the images captured by the participant.

After completing the interview the researchers coded and analyzed both the interview and the captured photos. They report their findings elsewhere [1]. We coded the interview and photo data as well, but concentrated on issues relating to the method itself. Below we present results both from our coding of the participant data as well as interviews with researchers about the method.

Results from participants

The median participant captured 34 photos (minimum participant: 15, maximum participant: 90). In the interviews, the researcher would progress through the photos in the order they were taken by the participant. Our analysis of the interviews and photos revealed that more time was spent with photographs viewed at the beginning of the interview than those viewed toward the end and that often participants and researchers would reference photos out-of-sequence. Also, participants would often photograph an object that itself was meaningless but that would cue recall of a specific event ("pointers"). We also found several cases in which participants adjusted their photo-taking process because of the presence of other people. Other findings included a prevalence of dynamic objects that do not lend themselves to photos, peripheral information in photos that was important in the elicitation interview and many instances of staged photos (*i.e.*, photos in which participants arranged objects or people specifically to make them easier to photograph).

We found that in every case, interviews followed themes rather than the temporal order of capture. Thus, most of the interview was spent on the first quartile of photos taken because the first photos would touch off a discussion about general habits that would not be revisited upon seeing later photos. For example, if someone got a cell phone call in the morning, a photo of that event would lead to a general discussion of cell phone use. In that case when photos of cell phones were viewed later in the interview, they generated far less discussion. Also, for some participants the interview would begin to concentrate on the topic of a later photo. In these cases, the researcher or participant would often "save that topic for later" and continue with the discussion with the intent of picking up that thread of conversation again when they arrived at the appropriate photo. Six out of 11 participants showed this trend, with total occurrences ranging from one to six times during their respective interviews.

Nine out of 11 participants used pictures of pointers, or objects that were not themselves information events but were reminders of other events, from three to 10 times during their interviews. These objects that served as pointers were usually related in a peripheral way to the event itself. For example, in one case a discussion with a colleague about an article read online was recalled via a picture of the beverage the colleague brought for the participant at the time. Also, there were five instances of pointers that were themselves direct side effects of the information event captured, such as a photo of diagrams on the wall as a reminder of the information conveyed during the meeting that produced them. Furthermore, each of the three times a pointer referred to a daily event (*e.g.*, getting the morning paper) the pointer led to only general recall of the event, with participants using such qualifiers as "usually" without uncovering the specifics of the instance. In every other case in which pointers were used they referred to unusual events and recall was specific.

In addition, we found that six out of 11 participants adjusted the style of capture because of the presence of people in the picture from two to seven times per interview. In one such case, a participant took a picture clandestinely to avoid fur-

ther aggravating an angry family member. Also, during the interviews the researchers conducted, there were two incidents of tangible objects serving important but unexpected roles as prompts. In one case in particular, the poor physical appearance of a participant's cell phone prompted follow-up questions about the participant's attitude toward the device.

Similar to others who have used the photo-elicitation method, we found that the most important information gleaned from the interview usually came not from the photos themselves but from the participant's description of the actions and processes that led them to take that image. However, we did note in seven different cases participants were unable to recall why they took an image. This was usually due to either the image resolving poorly or the capture of overloaded objects. For example, one participant took several pictures in a row of his cell phone. The elicitation session revealed that he had answered several phone calls in a row, but could not remember the specific content of those calls.

We also recorded the following findings:

- In all participant interviews there were situations in which peripheral information cued important discussions (*e.g.*, something not intentionally captured became a topic for discussion). The number of references to peripheral information ranged from one to 11 times per interview with a median of three occurrences. In nearly all of these cases, the researcher, not the participant, first referred to the peripheral information during the elicitation interview.
- In all participant interviews there were instances of staged photos, or those in which the participant arranged the scene or in which a person was photographed presenting for the camera. The number of staged photos ranged from two to 26 times per interview with a median of nine occurrences.
- Eight participants referenced other media in their photos. These were events in which participants took a picture of a physical artifact of some other medium, usually audio. Of the participants that showed this trend, the number of referenced media ranged from one to four times per interview with a median of 1.5 occurrences.
- Ten participants took photos that did not record correctly, usually because of lighting issues. Of the participants that showed this trend, the number of improper recordings ranged from one to five with a median of two. However, in nearly all cases the photos nonetheless led to recall of a specific event.
- Nine participants referenced some object in the interview that was itself never captured. Of the participants that showed this trend, the number of non-captured references ranged from one to nine times per interview with a median of four occurrences.

Results from researchers

From our interviews with researchers running the study we found a need for situated annotation as well as a means to review captured data and annotations before the elicitation interview takes place. The researchers commented that they felt they spent too much time on just a few images and having the chance to review and categorize them would facilitate their getting the most data out of limited interview time. However, they noted that in many cases the thing being photographed was not necessarily evident and thus some form of annotation of the photo would be crucial for them to categorize photos appropriately. They also commented that getting a few written responses to the photos would be helpful as well, but that it was unlikely that participants would complete such questions in the field. To that end, they expressed interest in a software tool to support desituated photo feedback, but were concerned about limiting their study population to people who have access to a computer. To remedy this problem, they brainstormed a public, community computer with photo feedback software that participants could easily access.

Transit decisions diary study

This study was run by another researcher at our institution. Similar to the first study, we were involved in the study from beginning to end, observed the process of using the method, analyzed results from the study and interviewed the researcher about his experience. This was a hybrid feedback and elicitation study that explored how people make public transit decisions. The researcher used phone-based feedback as well as location capture for elicitation. We analyzed feedback from participants as well as the results of the elicitation interview. We also interviewed the researcher about his experience with the method.

Method

The researcher provided four college students with a cell phone for a two week period and asked them to call a specified number every time they made a transit decision. When they called they were led through a series of questions about the event. Also, the location that they placed the call was automatically derived from a built-in GPS sensor and communicated to our server. The researcher then conducted two elicitation interviews: one a week into the study and the other immediately after the study was complete. During the first interview, the researcher used transcriptions of participant recordings as prompts, and during the second interview he used both the transcriptions as well as maps indicating the location of the participant when he or she completed a response.

Participants were asked the following questions each time they called: 1) Where are you going to and coming from? 2) How are you traveling? 3) What are you doing during your travel? 4) Do you expect to arrive early, on time or late? 5) How long do you expect to wait? 6) Did you consult any resources when you were planning this trip? 7) Is there anything special about this trip?

Results

As mentioned, feedback studies tend to place a heavy burden on participants because they require participants to switch tasks. However, the events about which participants provided feedback, transit decisions, occur relatively infrequently, reducing the burden on participants and yielding relatively high response rates.

Two participants logged responses a median of four times per day (Monday-Friday) while the other two responded a median of two times per day. The responses generally oc-

curred at the beginning and end of the day, corresponding to morning and evening commutes. However, in some of these cases participants responded only to correct a perceived mistake in an earlier response, and removing these repeat responses moves the median responses from four to three per day for one participant. Weekend response rates and times were much more sporadic, ranging from zero to two per day for one participant to one to five for another and with no specific pattern for any participant. The median time for individual participants to complete a set of answers per call ranged from one minute 37 seconds for one participant to two minutes one second for another.

An interview with the researcher revealed that, while referring to transcribed responses was helpful in recreating specific recording events during elicitation interviews, the maps neither aided participant recall nor were helpful to the researcher for logging purposes. Several reasons were given for this, including that location was not always captured for every event because of GPS coverage issues, that the maps could not dynamically show a sequence of calls and that the maps lacked detail. The researcher noted that if he had been able to visualize the sequence of responses it would have been easier to reconstruct the transit event, compare it to other possible events and ask the participant about their choices.

Festival diary study

To gain first-hand insight into the issues involved in running a diary study, we ran a study ourselves based on our first two studies and using the diary study method. During the study, we took notes on methodological breakdowns and concentrated on the method when coding data. From a sociological perspective, the study was designed to understand people's experience of novel information in non-everyday setting. From a methodological perspective, we compared the standard medium for elicitation studies, photography, with two other media, tangible objects and audio. As a contrast to the previous two studies, we chose to look at an unusual situation, a festival, rather than an everyday context.

Method

The focus of this study was to understand people's experience of novel information in non-everyday settings. We ran the study in a different context and with different media capture devices from the previous studies. We recruited seven college-educated participants, five women and two men, to capture information-related events they experienced during one day at a nearby jazz festival. We divided the participants into groups: two of the participants used digital cameras, two used digital audio recorders (Aigo mp3 player/recorders) and two were asked to collect tangible objects in a bag. Also, we asked one participant to capture both audio data and tangible objects.

After we provided tutorials on the use of the capture technology, participants captured information during one day of the festival and were interviewed about their items immediately afterwards. After completing the interviews we again coded the interviews and the captured media and we cataloged all of the objects collected by tangible-media participants. In this case we coded the data only for issues related to method, especially those that had to do with differences and similarities in capture media.

Results

During the one day that the participants participated in the study, the photo-elicitation participants collected 56 and 42 images, the audio participants collected 25 (median length: 1 minute and 32 seconds) and 45 (median length: 21 seconds) recordings, and the tangible object participants collected 28 and 14 distinct objects. Also, the seventh participant collected 12 recordings (median length: 1 minute and 3 seconds) and 13 objects.

Tangible objects

Of the 55 total tangible objects the three participants collected, 30 were information objects themselves (*e.g.*, flyers), 14 were pointers to some information event while 11 were side effects of some event, for example an extra copy of a form that a participant completed. Also, because this was an outdoor event, many of the objects collected were natural objects, such as a leaf from a tree, but represented entirely different events and media (the leaf was used to both represent music as well as an encounter with a friend). Also, the type of events prompted by the tangible objects varied considerably. In one case, a piece of bark prompted a participant to discuss in considerable detail a complicated event in which she helped a handicapped festival-goer physically maneuver in a crowded spot (see Figure 1, upper left). In another case, a participant grabbed three flowers to represent a general idea she had written in a personal diary during the festival.

Also, during the elicitation interview we found the participant's spatial arrangement of the items to be important. In the case mentioned above, the participant arranged the flowers on the table in a gradient from brightest to darkest and explained that the arrangement mapped her opinions of various flavors of jazz: Latin (bright), cool jazz (middle) and traditional (dark) (see Figure 1, upper left). Also, another participant arranged all of her items during the interview by narrative. For example, one of her narratives involved being asked to dance by a man, and she used both a bottle the man had with him and a jalapeño he gave to her as prompts during her explanation of the story (see Figure 1, upper right).

We noticed that memory of event order was poor. Two of the participants corrected their recall of the order of events three times and the other made four corrections. Also, two of the participants were not able to recall any details about the ambient audio at the time of capture, and the other participant recalled ambient audio roughly half of the time.

Audio recordings

The type of audio that was recorded varied significantly by participant. One participant captured almost entirely (23 out of 25) music events, while the other captured almost entirely (40 out of 45) ambient events, such as people talking, planes going by or overheard conversations. The other participant tended to capture more music events (8 out of 12).

We found that after identifying the contents of the recording, participants had excellent recall of the event. During

the interview we stepped through each of the recordings in the order that they were made, so all participants were able to recall the sequencing of events. Participants universally recalled the place the recording occurred. However, participant *identification* of the event took longer than anticipated (45 seconds of playing time on average). While recording quality was often the reason for poor recognition, in some cases the event of interest to the participant simply did not occur at the beginning of the recording. In those cases, participants tended to describe general features of the event ("This sounds like a bunch of people talking, and I was really interested in their voices") until some point in the recording when they were suddenly able to recall the specific event and then describe that event ("Oh, laughing, that's it, I was fascinated by how these two people were laughing...").

We found that participants used audio to clandestinely capture events that they otherwise may not have. One participant captured 12 events of other people talking, masking her recording by pretending to be doing something else, for example hiding the recording device behind a book or in her palm while looking another direction. In these cases we asked participants how they would have captured such data if they had a video device. In most cases the participants said they would not have recorded the event unless they asked permission first, in which case, as one participant said, "it would not have been very realistic."

Also, during elicitation, the participant who used both audio and tangible capture referred to some events being "linked," or audio that annotated a tangible object, and we spent some time searching for the appropriate audio recording before finding it.

Images

Participants captured 43 and 55 images. Unlike the everyday study, the interviews largely followed temporal order of capture more than the thematic order of capture, likely because most images were of one-time events. Also, there were only five total pointer events. That is, most of the photos contained information about the actual event that the participant experienced.

There were 37 total media cue events, but this was likely due to the fact that the music had a physical representation (the performers themselves). However, audio recall was poor when cues were not provided in the photos themselves.

Though participants rarely adjusted photos taken of another person out of privacy concerns, there were several other occasions in which the participant's picture-taking affected others. For example, one participant clandestinely took a picture of a man who was sketching a young woman sitting in the crowd, unbeknownst to the woman. The woman, seeing the participant take the image, began to take an interest in what the man was doing and ultimately the man offered his sketch to the woman.

Finally, peripheral information in the images was again significant, playing a role in 20 elicitations.

Comparing results between media

The results from the festival study allowed us to compare audio, photo and tangible object elicitation. As Barsalou argued, it is important to analyze to what extent each capture medium supports recognition of *who*, *where* and *what* information about the captured event [3]. In each case, participants were able to recall people involved in an event with whom they were already familiar. Photos provide the best support of *who* and *where* recognition, while audio clips, once recognized, also provide adequate support. Tangible objects did not lend themselves to *who* or *where* recognition.

Timing and sequencing of events are important for activity reconstruction. Participants tended to have poor recall of the exact time of capture for all media. Also, for tangible objects participants tended to be unable to recall the sequencing of events, while photo and audio capture are inherently sequenced.

In addition, participants generally were unable to recall information on media channels other than the cue. When a visual representation of the media was captured, it was often difficult or impossible to recognize and participants often could not recall the specific media. For example, a participant who took a picture of his audio-playing software could neither recognize (because of the fidelity of the picture) nor recall what music he was listening to. However, the photo did lead to a general discussion about audio consumption habits.

DISCUSSION

Results from the use of different media in diary studies suggest adjustments to the method to better accommodate different situations. Specifically, we found that audio elicitation suffers from recognition problems but encourages more clandestine capture events. Also, tangible objects are more likely than other media to elicit from participants creative explanations of attitudes and beliefs. The results also revealed the need for new tools to support the method. In particular, we found that tools are needed to support tagging of tangible objects; lightweight, situated annotation; researcher review of captured events; and automatic time stamping for all captured events.

We found that, overall, photos are the easiest to capture and recognize. However, audio cues can allow participants to capture events clandestinely that they otherwise may not have. Also, audio is a lightweight media appropriate for annotation. In general, for studies in which detail is important, a hybrid photo/audio capture medium is most appropriate.

From the transit study, we found that raw location information is not likely to lead to better recall of an episode. However, we also found that participants seem to be willing to spend longer answering feedback questions when the rate of events to report is low. These two issues suggest that situated feedback may be appropriate for some studies, but that feedback should be tied to better prompting cues. One way of supporting this would be to use photo-elicitation combined with more structured annotations, in which participants are encouraged to answer a set of specific questions.

Similar to location information, tangible objects are not likely to cue episodic memory. Thus, tangible objects are not appropriate for studies in which detailed recall is important. However, this lack of specificity could be a benefit for studies that concentrate less on the reconstruction of specific events and more on participant's attitudes and beliefs.

We found that similar to cultural probes, tangible object elicitation may inspire unique ways of describing and codifying beliefs and behavior [10]. Also, while cultural probes are able to inspire responses to general feelings about a community and culture, auto-driven object elicitation helps inspire recall and description of specific events. We also found that it would be useful to tag objects with audio annotations.

To support recall of ambiguous events, we found that it is important that each capture event be tagged with a brief annotation. However, as noted above, in the photo-elicitation study pilot participants who were asked to annotate their pictures with written diaries usually gave up the practice immediately as it was too disruptive. Thus, rapid, situated annotation, such as audio for photographs and tangible objects, is crucial. We also found that researchers need to be able to review captured data as well as annotations before the elicitation takes place. This preparation is necessary to get the most out of limited interview time as possible. Because peripheral information in photographs consistently provided useful information, a tool that allows participants and researchers to annotate various parts of the photographs would be useful. In general, the results of our studies argue for lightweight capture tools combined with lightweight *in situ* annotation as well as support for more thorough *ex situ* annotation and review by both participants and researchers.

Our studies and others indicate that people are not good at judging how long an activity takes to complete [16]. Thus, automatic time stamping as often as possible is crucial. Our experience also revealed that in studies of everyday situations that depend on empirical evidence, rather than general participant attitudes, researchers should only encourage the capture of "pointers" to events when those events are unusual. Capture of regularly experienced events should be direct to avoid generalization.

A PROPOSED DIARY STUDY PIPELINE

As we mentioned earlier, feedback diary studies and elicitation diary studies represent a tradeoff between accurate recall but burdensome logging (feedback) versus potentially inaccurate recall but unobtrusive logging (elicitation). In our experience with media-based diary studies as well as reports in the literature, we found that it is important to mitigate the impact of a study on participant's everyday interactions and encourage participant recall of ambiguous data. We also found that for elicitation studies it is important to provide support for interview preparation. To address these issues we propose a tool-supported diary study pipeline that borrows from both feedback and elicitation methods to maximize participant recall and interview preparation while minimizing situated logging. Specifically, the pipeline includes (1) lightweight *in situ* capture by participants augmented with (2) lightweight *in situ* annotation at the time of capture

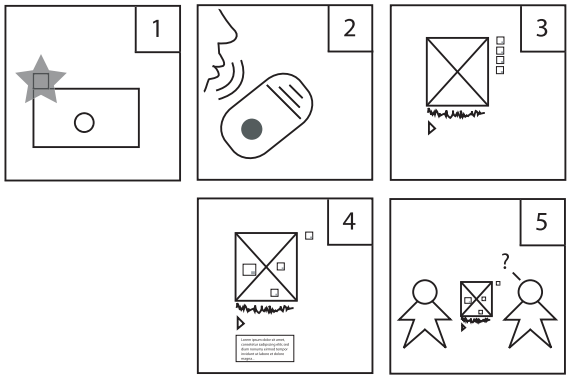

Figure 2. Proposed media elicitation pipeline: 1) A participant takes a photo 2) The participant annotates the photo with an audio recording 3) The participant uses a tool to log the photo and audio and add more annotations 4) The researcher provides feedback about the captured data 5) The researcher holds an elicitation interview with the participant using the captured media as prompts.

to encourage recall, followed by (3) more extensive annotation by participants at a later time, allowing for (4) review of the data by researchers to better structure (5) a post-study interview (see Figure 2). This pipeline minimizes the extent to which participants are distracted from their primary tasks while still allowing them to recall and comment on the event at a more convenient time. Furthermore, unlike any previously conducted media-based diary study, researchers have the opportunity to prepare for elicitation interviews based on specific data.

Reporter

As a first attempt to support the pipeline described above, we present Reporter. We hope that this tool can aid researchers performing diary studies that use capture technologies, but it is not a replacement for other tools and methods necessary to conduct a study. Reporter combines a lightweight Java client and a Web interface to facilitate diary studies that involve digital capture media. Specifically, Reporter allows researchers to configure per-capture questions and consolidates and provides support for capture annotations. The steps in a study that would use Reporter are approximately the following: 1) A researcher enters questions that participants will answer about each piece of captured data using Reporter's Web interface 2) Participants download and install a small Java client to a desktop machine 3) Participants capture events and audio annotations in the field as per the researchers instructions during some period of time 4) When the participant is able to return to her desktop she uses the Reporter client to upload the data she has collected and then uses a Web interface launched from the client to answer per-capture questions 5) The researcher uses the responses and photo data to structure a subsequent post-study interview.

If the study occurs over the course of several days, the researcher may use the Web interface to provide feedback or ask follow-up questions of participants about specific photos. After the researcher attaches a feedback question to a

photo, the Web interface flags that photo so that participants can rapidly review all of the outstanding follow-up questions they have yet to answer. Also, a researcher may use the tool completely asynchronously (*i.e.*, as a feedback tool rather than an elicitation tool).

Our design includes support for two types of photo annotation. First, users may use the Java client to upload digital audio clips captured in the field. Second, we provide a way for the researcher to ask participants to tag parts of photos using semi-transparent rectangles that are movable and scalable. The photo annotation rectangles are implemented in DHTML and thus do not significantly impact the load-time of the page nor require any specific plug-ins. Also, clicking on the rectangles rotates through a series of colors, allowing participants to group annotations by color. As an example, consider Figure 3. Here a participant has used the photo annotation tool to designate features requested by the researcher: the object that the participant actually intended to capture (upper left of the photo) and any other information that the participant felt was important in the scene (center and right of the photo). The participant has also uploaded an audio annotation captured in the field (lower left of the figure).

Other tools have been built to support field studies, but none can be used to support the pipeline above. Intille *et al.* designed a system for photo-based ESM, but it relies upon an infrastructure that would be difficult to implement in everyday field settings and involves context-aware capture rather than participant capture [14]. Cybertracker and systems developed by Pascoe *et al.* allow field workers to track animals by providing a memory prosthesis, but those systems neither require nor provide support for interactions between researchers and participants [8, 20]. Also, others have developed episodic memory prompts but that were not used for evaluation. Forget-me-not automatically captured event context and displayed icon-based cues for each event [17]. Eldridge *et al.* used video to aid recall but again did not use the system for evaluation purposes [9]. Also, Carmien explored prompts on personal devices for personal coaches for the memory impaired [5].

User tests

We pilot tested the five-step pipeline with two participants. In this study, participants captured information production and consumption events for one day and used Reporter to upload pictures and audio annotations, visually annotate photos, and answer a few questions about each event. Participants used digital cameras with audio annotation features to record events. We asked participants the following questions: 1) Please indicate the area that you were trying to take a picture of with a red square and describe what or who it is (in the textbox) below. 2) If there is anything in this photograph that you want to label and comment on, please use yellow colored squares and discuss them (in the textbox) below. 3) Did you talk to any people in this photograph? If so, use blue squares to designate them and indicate who they are (in the textbox) below (not names, just how they relate to you). If not, write "none." 4) Please discuss briefly how important this object/person is to you. 5) How often do use this object/person as an information resource (only once/hourly/daily/weekly/monthly/yearly)?

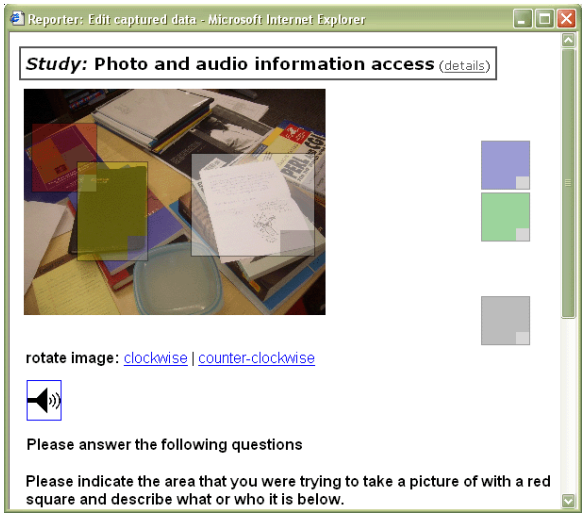

Figure 3. Reporter's Web interface. The participant has repositioned and resized rectangles as per questions asked by researchers, including the object that the participant intended to capture (upper left of the photo) and any other information that the participant felt was important in the scene (center and right of the photo).

The day after we reviewed the data and used Reporter's feedback feature to ask follow-up questions about certain photos. The participants then answered the questions using Reporter's Web interface and we then interviewed them about the photos they captured as well as their experience using Reporter.

The participants found the system easy to learn. In particular, participants mastered the visual photo annotation technique within minutes. Also, we found as researchers that we required a means to add links to specific study instructions at pertinent parts of the interface. For example, when uploading photos participants need to know the policy on photo deletion and modification, an issue which is likely to change per experiment.

We also found that audio recording is a mostly unobtrusive means of annotating. In other cases, participants took a photo and simultaneously cued the audio recording while continuing with the task in which they were involved. For example, one participant snapped a photo while walking and carrying on a conversation with a friend. The participant then continued walking and conversing while cueing the audio recording and used a break in the conversation to record the annotation far from the point of capture. Other types of media, such as video, would likely not have been able to support this kind of use, instead requiring the participant to stop what they are doing completely in order to capture and annotate the event. On the other hand, one participant commented that he was uncomfortable recording audio annotations in some locations, such as lecture halls.

Because we were able to view photos and annotations beforehand, we more effectively structured our elicitation interviews. Though we still found it useful to understand the sequence of events that participants captured, we organized the interview around themes. In some cases, the same photograph was important for different themes. For example, a picture of a computer also included peripheral audio cues, and we discussed each at different points in the interview.

CONCLUSION AND FUTURE WORK

In conclusion, we derived improvements to the diary study technique from three studies of the technique itself in action. Our studies revealed a need for situated annotation of captured event in elicitation studies. We also found that the best approach to feedback studies may be to pair media capture with structured, question-and-answer based annotations. Our studies also revealed the usefulness of different media in different situations.

We also developed and tested a diary study pipeline that borrows from both feedback and elicitation methods to maximize participant recall and interview preparation while minimizing situated logging. We then built and tested a lightweight tool, *Reporter*, to support this pipeline.

In future work, we intend to explore the use of other media in diary studies. In particular we are interested in exploring video, which has been shown aid recall [9]. We also intend to build and test prototypes for tagging tangible objects with audio annotations as well as for structured, question-based annotations.

ACKNOWLEDGMENTS

We thank Judd Antin, Joseph Lorenzo Hall, Dr. Peter Lyman, and Christopher Beckmann. This work was supported by NSF grants IIS-0209213 and IIS-0205644.

REFERENCES

1. J. Antin, J. L. Hall, and P. Lyman. Information filtering behavior in everyday life. *In submission*, 2005.
2. L. F. Barrett and D. J. Barrett. An introduction to computerized experience sampling in psychology. *Soc. Sci. Comput. Rev.*, 19(2):175–185, 2001.
3. L. W. Barsalou. The content and organization of autobiographical memories. In U. N. . E. Winograd, editor, *Remembering reconsidered: Ecological and traditional approaches to the study of memory*, pages 193–243. Cambridge University Press, 1988.
4. B. A. T. Brown, A. J. Sellen, and K. P. O'Hara. A diary study of information capture in working life. In *ACM CHI '00*, pages 438–445, 2000.
5. S. Carmien, R. DePaula, A. Gorman, and A. Kintsch. Increasing workplace independence for people with cognitive disabilities by leveraging distributed cognition among caregivers and clients. In *GROUP '03*, pages 95–104, 2003.
6. M. Clark-Ibáñez. Framing the social world with photo-elicitation interviews. *American Behavioral Scientist*, 47(12):1507–1527, 2004.
7. M. Colbert. A diary study of rendezvousing: implications for position-aware computing and communications for the general public. In *GROUP '01*, pages 15–23, 2001.
8. Cybertracker. http://www.cybertracker.co.za/.
9. M. Eldridge, M. Lamming, and M. Flynn. Does a video diary help recall? In *the HCI '92*, pages 257–269, 1992.
10. B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural probes. *interactions*, 6(1):21–29, 1999.
11. R. Grinter and M. Eldridge. Wan2tlk?: everyday text messaging. In *ACM CHI '03*, pages 441–448, 2003.
12. T. Hemmings, A. Crabtree, T. Rodden, K. Clarke, and M. Rouncefield. Probing the probes. In *Participatory Design Conference*, pages 42–50, 2002.
13. H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderback. Technology probes: inspiring design for and with families. In *ACM CHI '03*, pages 17–24, 2003.
14. S. Intille, C. Kukla, and X. Ma. Eliciting user preferences using image-based experience sampling and reflection. In *ACM CHI '02*, pages 738–739, 2002.
15. S. S. Intille, J. Rondoni, C. Kukla, I. Ancona, and L. Bao. A context-aware experience sampling tool. In *ACM CHI '03*, pages 972–973, 2003.
16. F. T. Juster. Conceptual and methodological issues involved in the measurement of time use. In F. T. Juster and F. P. Stafford, editors, *Time, Goods, and Well-Being*, pages 19–32. Institute for Social Research, The University of Michigan, 1985.
17. M. Lamming and M. Flynn. Forget-me-not: intimate computing in support of human memory. In *Proceedings FRIEND21*, 1994.
18. K. O'Hara and M. Perry. Shopping anytime anywhere. In *ACM CHI '01*, pages 345–346, 2001.
19. L. Palen and M. Salzman. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *ACM CSCW '02*, pages 87–95, 2002.
20. J. Pascoe, N. Ryan, and D. Morse. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*, 7(3):417–437, 2000.
21. M. G. Petersen, K. H. Madsen, and A. Kjaer. The usability of everyday technology: emerging and fading opportunities. *ACM Transactions on Computer-Human Interaction*, 9(2):74–105, 2002.
22. J. Rieman. The diary study: A workplace-oriented research tool to guide laboratory efforts. In *ACM INTERCHI '93*, pages 321–326, 1993.
23. A. V. Sampson-Cordle. *Exploring the relationship between a small rural school in Northeast Geogia and its community: an image-based study using participant-produced photographs*. PhD thesis, University of Georgia, 2001.
24. J. Samuels. Breaking the ethnographer's frames: Reflections on the use of photo elicitation in understanding sri lankan monastic culture. *American Behavioral Scientist*, 47(12):1528–1550, 2004.
25. A. Sellen and R. Harper. Paper as an analytic resource for the design of new technologies. In *ACM CHI 97*, pages 319–326, 1997.