8.6 Observation

Observation is useful at any stage during product development. Early in design, observation helps designers understand the users' context, tasks, and goals. Observation conducted later in development, for example, in evaluation, may be used to investigate how well a prototype supports these tasks and goals.

Users may be observed directly by the investigator as they perform their activities or indirectly through records of the activity that are studied afterward (Bernard, 2017). Observation may also take place in the field or in a controlled environment. In the former case, individuals are observed as they go about their day-to-day tasks in the natural setting. In the latter case, individuals are observed performing specified tasks within a controlled environment such as a usability laboratory.

ACTIVITY 8.5

To appreciate the different merits of observation in the field and observation in a controlled environment, read the following scenarios and answer the questions that appear after.

Scenario 1 A usability consultant joins a group of tourists who have been given a wearable navigation device that fits onto a wrist strap to test on a visit to Stockholm. After sightseeing for the day, they use the device to find a list of restaurants within 2 kilometers of their current position. Several are listed, and they find the phone numbers of a few, call them to ask about their menus, select one, make a booking, and head off to the restaurant. The usability consultant observes some difficulty operating the device, especially on the move. Discussion with the group supports the evaluator's impression that there are problems with the interface, but on balance the device is useful, and the group is pleased to get a table at a good restaurant nearby.

Scenario 2 A usability consultant observes how participants perform a preplanned task using the wearable navigation device in a usability laboratory. The task requires the participants to find the phone number of a restaurant called Matisse. It takes them several minutes to do this, and they appear to have problems. The video recording and interaction log suggest that the interface is quirky and the audio interaction is of poor quality. This is supported by participants' answers on a user satisfaction questionnaire.

1. What are the advantages and disadvantages of these two types of observation?
2. When might each type of observation be useful?

Comment

1. The advantages of the field study are that the observer saw how the device could be used in a real situation to solve a real problem. They experienced the delight expressed with the overall concept and the frustration with the interface. By watching how the group used the device on the move, they gained an understanding of what the participants liked and what was lacking. The disadvantage is that the observer was an insider in the group,

(Continued)

so how objective could they be? The data is qualitative, and while anecdotes can be very persuasive, how useful are they? Maybe they were having such a good time that their judgment was clouded and they missed hearing negative comments and didn't notice some of the participant's annoyance. Another study could be done to find out more, but it is not possible to replicate the exact conditions of this study. The advantages of the lab study are that it is easier to replicate, so several users could perform the same task, specific usability problems can be identified, users' performance can be compared, and averages for such measures as the time it took to do a specific task and the number of errors can be calculated. The observer could also be more objective as an outsider. The disadvantage is that the study is artificial and says nothing about how the device would be used in the real environment.

2. Both types of study have merits. Which is better depends on the goals of the study. The lab study is useful for examining details of the interaction style to make sure that usability problems with the interface and button design are diagnosed and corrected. The field study reveals how the navigation device is used in a real-world context and how it integrates with or changes users' behavior. Without this study, it is possible that developers might not have discovered the enthusiasm for the device because the reward for doing laboratory tasks is not as compelling as a good meal! In fact, according to Kjeldskov and Skov (2014), there is no definitive answer to which kind of study is preferable for mobile devices. They suggest that the real question is when and how to engage with longitudinal field studies. ■

8.6.1 Direct Observation in the Field

It can be difficult for people to explain what they do or to describe accurately how they achieve a task. It is unlikely that an interaction designer will get a full and true story using interviews or questionnaires. Observation in the field can help fill in details about how users behave and use technology, and nuances that are not elicited from other forms of investigation may be observed. Understanding the context provides important information about why activities happen the way that they do. However, observation in the field can be complicated and harder to do well than at first appreciated. Observation can also result in a lot of data, some of which may be tedious to analyze and not very relevant.

All data gathering should have a clearly stated goal, but it is particularly important to have a focus for an observation session because there is always so much going on. On the other hand, it is also important to be prepared to change the plan if circumstances change. For example, the plan may be to spend one day observing an individual performing a task, but an unexpected meeting crops up, which is relevant to the observation goal and so it makes sense to attend the meeting instead. In observation, there is a careful balance between being guided by goals and being open to modifying, shaping, or refocusing the study as more is learned about the situation. Being able to keep this balance is a skill that develops with experience.

Structuring Frameworks for Observation in the Field

During an observation, events can be complex and rapidly changing. There is a lot for observers to think about, so many experts have a framework to structure and focus their observation. The framework can be quite simple. For example, this is a practitioner's framework for use in evaluation studies that focuses on just three easy-to-remember items:

The person: Who is using the technology at any particular time?
The place: Where are they using it?
The thing: What are they doing with it?

Even a simple framework such as this one based on who, where, and what can be surprisingly effective to help observers keep their goals and questions in sight. Experienced observers may prefer a more detailed framework, such as the following (Robson and McCarten, 2016, p. 328), which encourages them to pay greater attention to the context of the activity:

Space: What is the physical space like, and how is it laid out?
Actors: What are the names and relevant details of the people involved?
Activities: What are the actors doing, and why?
Objects: What physical objects are present, such as furniture?
Acts: What are specific individual actions?
Events: Is what you observe part of a special event?
Time: What is the sequence of events?
Goals: What are the actors trying to accomplish?
Feelings: What is the mood of the group and of individuals?

This framework was devised for any type of observation, so when used in the context of interaction design, it might need to be modified slightly. For example, if the focus is going to be on how some technology is used, the framework could be modified to ask the following:

Objects: What physical objects, in addition to the technology being studied, are present, and do they impact on the technology use?

Both of these frameworks are relatively general and could be used in many different types of study, or as a basis for developing a new framework for a specific study.

ACTIVITY 8.6

1. Find a small group of people who are using any kind of technology, for example, smartphones, household appliances, or video game systems, and try to answer the question, "What are these people doing?" Watch for three to five minutes, and write down what you observe. When finished, note down how it felt to be doing this and any reactions in the group of people observed.
2. If you were to observe the group again, what would you do differently?
3. Observe this group again for about 10 minutes using the detailed framework given above.

(Continued)

Comment

1. What problems did this exercise highlight? Was it hard to watch everything and remember what happened? How did the people being watched feel? Did they know they were being watched? Perhaps some of them objected and walked away. If you didn't tell them that they were being watched, should you have?
2. The initial goal of the observation, that is, to find out what the people are doing, was vague, and chances are that it was quite a frustrating experience not knowing what was significant and what could be ignored. The questions used to guide observation need to be more focused. For example, you might ask the following: What are the people doing with the technology? Is everyone in the group using it? Are they looking pleased, frustrated, serious, happy? Does the technology appear to be central to the users' goals?
3. Ideally, you will have felt more confident this second time, partly because it is the second time doing some observation and partly because the framework provided a structure for what to look at. ■

Degree of Participation

Depending on the type of study, the degree of participation within the study environment varies across a spectrum, which can be characterized as insider at one end and outsider at the other. Where a particular study falls along this spectrum depends on its goal and on the practical and ethical issues that constrain and shape it.

An observer who adopts an approach right at the outsider end of the spectrum is called a *passive observer*, and they will not take any part in the study environment at all. It is difficult to be a truly passive observer in the field, simply because it's not possible to avoid interacting with the activities. Passive observation is more appropriate in lab studies.

An observer who adopts an approach at the insider end of this spectrum is called a *participant observer*. This means that they attempt, at various levels depending on the type of study, to become a member of the group being studied. This can be a difficult role to play since being an observer also requires a certain level of detachment, while being a participant assumes a different role. As a participant observer, it is important to keep the two roles clear and separate so that observation notes are objective while participation is also maintained. It may not be possible to take a full participant observer approach for other reasons. For example, the observer may not be skilled enough in the task at hand, the organization/group may not be prepared for an outsider to take part in their activities, or the timescale may not provide sufficient opportunity to become familiar enough with the task to participate fully. Similarly, if observing activity in a private place such as the home, full participation would be difficult even if, as suggested by some researchers (for example, Bell et al., 2005), you have spent time getting to know the family before starting the study. Chandrika Cycil et al. (2013) overcame this issue in their study of in-car conversations between parents and children by traveling with the families initially for a week and then asking family members to video relevant episodes of activity. In this way, they had gained an understanding of the context and family dynamics and then collected more detailed data to study activity in depth.

Planning and Conducting an Observation in the Field

The frameworks introduced in the previous section are useful for providing focus and also for organizing the observation and data gathering activity. Choosing a framework is important, but there are other decisions that need to be made, including the level of participation to adopt, how to make a record of the data, how to gain acceptance in the group being studied, how to handle sensitive issues such as cultural differences or access to private spaces, and how to ensure that the study uses different perspectives (people, activities, job roles, and so forth).

One way to achieve this last point is to work as a team. This can have several benefits.

- Each person can agree to focus on different people or different parts of the context, thereby covering more ground.
- Observation and reflection can be interweaved more easily when there is more than one observer.
- More reliable data is likely to be generated because observations can be compared.
- Results will reflect different perspectives.

Once in the throes of an observation, there are other issues that need to be considered. For example, it will be easier to relate to some people more than others. Although it will be tempting to pay attention to them more than others, attention needs to be paid to everyone in the group. Observation is a fluid activity, and the study will need to be refocused as it progresses in response to what is learned. Having observed for a while, interesting phenomena that seem relevant will start to emerge. Gradually, ideas will sharpen into questions that guide further observation.

Observing is also an intense and tiring activity, but checking notes and records and reviewing observations and experiences at the end of each day is important. If this is not done, then valuable information may be lost as the next day's events override the previous day's findings. Writing a diary or private blog is one way of achieving this. Any documents or other artifacts that are collected or copied (such as minutes of a meeting or discussion items) can be annotated, describing how they are used during the observed activity. Where an observation lasts several days or weeks, time can be taken out of each day to go through notes and other records.

As notes are reviewed, separate personal opinion from observation and mark issues for further investigation. It is also a good idea to check observations and interpretations with an informant or members of the participant group for accuracy.

DILEMMA

When to Stop Observing?

Knowing when to stop doing any type of data gathering can be difficult for novices, but it is particularly tricky in observational studies because there is no obvious ending. Schedules often dictate when your study ends. Otherwise, stop when nothing new is emerging. Two indications of having done enough are when similar patterns of behavior are being seen and when all of the main stakeholder groups have been observed and a good understanding of their perspectives has been achieved. ■

Ethnography

Ethnography has traditionally been used in the social sciences to uncover the organization of societies and their activities. Since the early 1990s, it has gained credibility in interaction design, and particularly in the design of collaborative systems; see Box 8.6, "Ethnography in Requirements" and Crabtree (2003). A large part of most ethnographic studies is direct observation, but interviews, questionnaires, and studying artifacts used in the activities also feature in many ethnographic studies. A distinguishing feature of ethnographic studies compared with other data gathering is that a situation is observed without imposing any *a priori* structure or framework upon it, and everything is viewed as "strange." In this way, the aim is to capture and articulate the participants' perspective of the situation under study.

BOX 8.6

Ethnography in Requirements

The MERboard is a tool scientists and engineers use to display, capture, annotate, and share information in support of the operation of two Mars Exploration Rovers (MERs) on the surface of Mars. The MER (see Figure 8.8) acts like a human geological explorer by collecting and analyzing samples and then transmitting the results to the scientists on Earth. The scientists and engineers collaboratively analyze the data received, decide what to study next, create plans of action, and send commands to the robots on the surface of Mars.

Figure 8.8 Mars Exploration Rover
Source: NASA Jet Propulsion Laboratory (NASA-JPL)

The requirements for MERboard were identified partly through ethnographic fieldwork, observations, and analysis (Trimble et al., 2002). The team of scientists and engineers ran a series of field tests that simulated the process of receiving data, analyzing it, creating plans, and transmitting them to the MERs. The main problems they identified stemmed from the scientists' limitations in displaying, sharing, and storing information (see Figure 8.9a).

(a) (b)

Figure 8.9 (a) The situation before MERboard; (b) a scientist using MERboard to present information

Source: Trimble et al. (2002)

These observations led to the development of MERboard (see Figure 8.9b), which contains four core applications: a whiteboard for brainstorming and sketching, a browser for displaying information from the web, the capability to display personal information and information across several screens, and a file storage space linked specifically to MERboard. ■

Ethnography has become popular within interaction design because it allows designers to obtain a detailed and nuanced understanding of people's behavior and the use of technology that cannot be obtained by other methods of data gathering (Lazar et al., 2017). While there has been much discussion of how big data can address many design issues, big data is likely to be most powerful when combined with ethnography to explain how and why people do what they do (Churchill, 2018).

The observer in an ethnographic study adopts a participant observer (insider) role as much as possible (Fetterman, 2010). While participant observation is a hallmark of ethnographic studies, it is also used within other methodological frameworks such as action research (Hayes, 2011), where one of the goals is to improve the current situation.

Ethnographic data is based on what is available, what is "ordinary," what it is that people do, say, and how they work. The data collected therefore has many forms: documents, notes taken by the observer(s), pictures, and room layout sketches. Notes may include snippets of conversations and descriptions of rooms, meetings, what someone did, or how people reacted to a situation. Data gathering is opportunistic, and observers make the most of opportunities as they present themselves. Often, interesting phenomena do not reveal themselves immediately but only later, so it is important to gather as much as possible within the framework of observation. Initially, spend time getting to know people in the participant group and bonding with them. Participants need to understand why the observers are there, what they hope

to achieve, and how long they plan to be there. Going to lunch with them, buying coffee, and bringing small gifts, for example, cookies, can greatly help this socialization process. Moreover, key information may be revealed during one of these informal gatherings.

It is important to show interest in the stories, gripes, and explanations that are provided and to be prepared to step back if a participant's phone rings or someone else enters the workspace. A good tactic is to explain to one of the participants during a quiet moment what you think is happening and then let them correct any misunderstandings. However, asking too many questions, taking pictures of everything, showing off your knowledge, and getting in their way can be very off-putting. Putting up cameras on tripods on the first day may not be a good idea. Listening and watching while sitting on the sidelines and occasionally asking questions is a better approach.

The following is an illustrative list of materials that might be recorded and collected during an ethnographic study (adapted from Crabtree, 2003, p. 53):

- Activity or job descriptions
- Rules and procedures (and so on) that govern particular activities
- Descriptions of activities observed
- Recordings of the talk taking place between parties involved in observed activities
- Informal interviews with participants explaining the detail of observed activities
- Diagrams of the physical layout, including the position of artifacts
- Photographs of artifacts (documents, diagrams, forms, computers, and so on) used in the course of observed activities
- Videos of artifacts as used in the course of observed activities
- Descriptions of artifacts used in the course of observed activities
- Workflow diagrams showing the sequential order of tasks involved in observed activities
- Process maps showing connections between activities

Traditionally, ethnographic studies in this field aim to understand what people do and how they organize action and interaction within a particular context of interest to designers. However, recently there has been a trend toward studies that draw more on ethnography's anthropological roots and the study of culture. This trend has been brought about by the perceived need to use different approaches because the computers and other digital technologies, especially mobile devices, are embedded in everyday activity, and not just in the workplace as in the 1990s.

BOX 8.7

Doing Ethnography Online

As collaboration and social activity online have increased, ethnographers have adapted their approach to study social media and the various forms of computer-mediated communication (Rotman et al., 2013; Bauwens and Genoud, 2014). This practice has various names, the most common of which are online ethnography (Rotman et al., 2012), virtual ethnography (Hine, 2000), and netnography (Kozinets, 2010). Where a community or activity has both an online and offline presence, it is common to incorporate both online and offline techniques within the data gathering program. However, where the community or activities of interest exist almost exclusively online, then mostly online techniques are used and virtual ethnography becomes central.

Why is it necessary to distinguish between online and face-to-face ethnography? It is important because interaction online is different from interaction in person. For example, communication in person is richer (through gesture, facial expression, tone of voice, and so on) than online communication, and anonymity is more easily achieved when communicating online. In addition, virtual worlds have a persistence, due to regular archiving, that does not typically occur in face-to-face situations. This makes characteristics of the communication different, which often includes how ethnographers introduce themselves to the community, how they act within the community, and how they report their findings. For these reasons, some researchers who work primarily online also try to meet with some of the participants face-to-face, particularly when working on sensitive topics (Lingel, 2012).

Special tools may be developed to support ethnographic data collection. Mobilab is an online collaborative platform that was developed for citizens living in Switzerland to report and discuss their daily mobility during an eight-week period using their mobile phones, tablets, and computers (Bauwens and Genoud, 2014). Mobilab enabled the researchers to more easily engage in discussion with participants on a variety of topics, including trucks parking on a bikeway.

For observational studies in large social spaces, such as digital libraries or Facebook, there are different ethical issues to consider. For example, it is unrealistic to ask everyone using a digital library to sign any kind of form agreeing to be involved in the study, yet participants do need to understand the observer's role and the purpose of their study. The presentation of results needs to be modified too. Quotes from participants in the community, even if anonymized in the report, can easily be attributed by a simple search of the community archive or the IP address of the sender, so care is needed to protect their privacy. ■

8.6.2 Direct Observation in Controlled Environments

Observing users in a controlled environment may occur within a purposely built usability lab, but portable labs that can be set up in any room are quite common. Portable laboratories can mean that more participants take part because they don't have to travel away from their normal environment. Observation in a controlled environment inevitably takes on a more formal character than observation in the field, and the user may feel more apprehensive. As with interviews, it is a good idea to prepare a script to guide how the participants will be greeted, be told about the goals of the study and how long it will last, and have their rights explained. Use of a script ensures that each participant will be treated in the same way, which brings more credibility to the results obtained from the study.

The same basic data recording techniques are used for direct observation in the laboratory and field studies (that is, capturing photographs, taking notes, collecting video, and so on), but the way in which these techniques are used is different. In the lab the emphasis is on the details of what individuals do, while in the field the context is important, and the focus is on how people interact with each other, the technology, and their environment.

The arrangement of equipment with respect to the participant is important in a controlled study because details of the person's activity need to be captured. For example, one camera might record facial expressions, another might focus on mouse and keyboard activity, and another might record a broad view of the participant and capture body language. The stream of data from the cameras can be fed into a video editing and analysis suite where it is coordinated and time-stamped, annotated, and partially edited.

The Think-Aloud Technique

One of the problems with observation is that the observer doesn't know what users are thinking and can only guess from what they see. Observation in the field should not be intrusive, as this will disturb the context the study is trying to capture. This limits the questions being asked of the participant. However, in a controlled environment, the observer can afford to be a little more intrusive. The think-aloud technique is a useful way of understanding what is going on in a person's head.

Imagine observing someone who has been asked to evaluate the interface of the web search engine Lycos.com. The user, who does not have much experience of web searches, is told to look for a phone for a 10-year-old child. They are told to type www.lycos.com and then proceed however they think best. They type the URL and get a screen similar to the one in Figure 8.10.

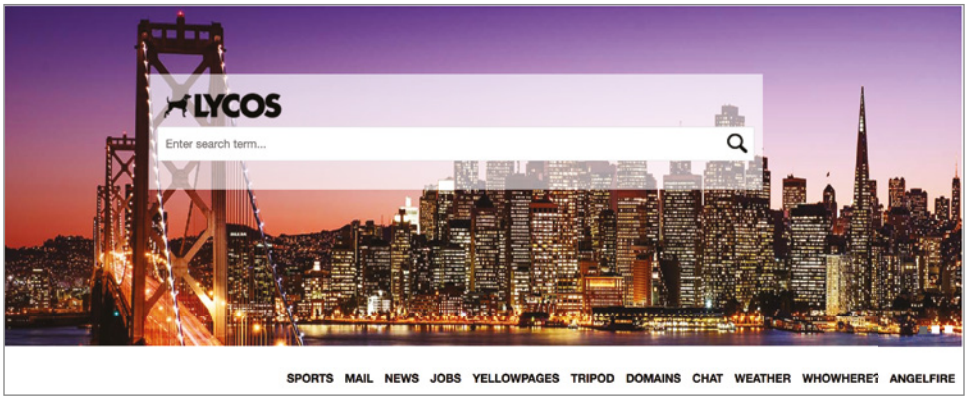

Figure 8.10 Home page of Lycos search engine
Source: https://www.lycos.com

Next, they type **child's phone** in the search box. They get a screen similar to the one shown in Figure 8.11. They are silent. What is going on? What are they thinking? One way around the problem of knowing what they are doing is to collect a think-aloud protocol, a technique developed by Anders Ericsson and Herbert Simon (1985) for examining people's problem-solving strategies. The technique requires people to say out loud everything that they are thinking and trying to do so that their thought processes are externalized.

So, let's imagine an action replay of the situation just described, as follows, but this time the user has been instructed to think aloud:

> *"I'm typing in www.lycos.com, as you told me." <types>*
> *"Now I am typing* **child's phone** *and then clicking the search button.*
> *<pause and silence>*
> *"It's taking a few seconds to respond."*
> *"Oh! Now I have a choice of other websites to go to. Hmm, I wonder which one I should select. Well, it's for a young child so I want a 'child-safe phone.' This one mentions safe phones <He clicks on 7 Best Cell Phones for Kids - Mashable>*
> *"Gosh, there's a lot more models to select from than I expected! Hmm, some of these are for older children. I wonder what I do next to find one for a 10-year-old."*

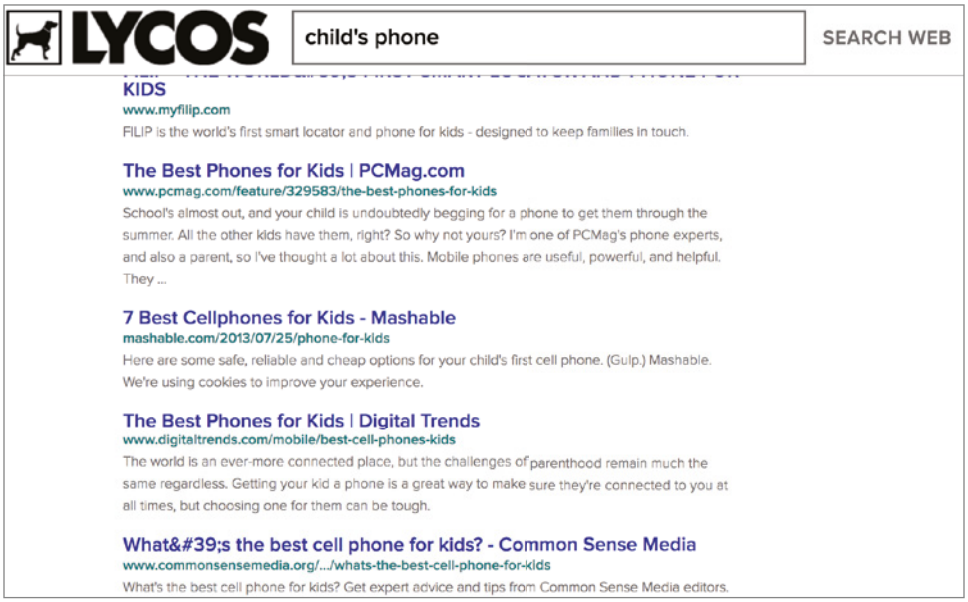

Figure 8.11 The screen that appears in response to searching for "child's phone"
Source: https://www.lycos.com

<pauses and looks at the screen>
I guess I should scroll through them and identify those that might be appropriate."
<silence . . . >

Now you know more about what the user is trying to achieve, but they are silent again. They are looking at the screen, but what are they thinking now? What are they looking at?

The occurrence of these silences is one of the biggest problems with the think-aloud technique.

ACTIVITY 8.7

Try a think-aloud exercise yourself. Go to a website, such as Amazon or eBay, and look for something to buy. Think aloud as you search and notice how you feel and behave.

Afterward, reflect on the experience. Was it difficult to keep speaking all the way through the task? Did you feel awkward? Did you stop talking when you got stuck?

Comment

Feeling self-conscious and awkward doing this is a common response, and some people say they feel really embarrassed. Many people forget to speak out loud and find it difficult to do so when the task becomes difficult. In fact, you probably stopped speaking when the task became demanding, and that is exactly the time when an observer is most eager to hear what's happening.

(Continued)

If a user is silent during a think-aloud protocol, the observer could interrupt and remind them to think out loud, but that would be intrusive. Another solution is to have two people work together so that they talk to each other. Working with another person (called *constructive interaction* [Miyake, 1986]) is often more natural and revealing because participants talk in order to help each other along. This technique has proved to be particularly successful with children, and it also avoids possible cultural influences on concurrent verbalization (Clemmensen et al., 2008). ■

8.6.3 Indirect Observation: Tracking Users' Activities

Sometimes direct observation is not possible because it is too intrusive or observers cannot be present over the duration of the study, and so activities are tracked indirectly. Diaries and interaction logs are two techniques for doing this.

Diaries

Participants are asked to write a diary of their activities on a regular basis, including things like what they did, when they did it, what they found hard or easy, and what their reactions were to the situation. For example, Sohn et al. (2008) asked 20 participants to record their mobile information needs through text messages and then to use these messages as prompts to help them answer six questions on a website at the end of each day. From the data collected, they identified 16 categories of mobile information needs, the most frequent of which was "trivia."

Diaries are useful: when participants are scattered and unreachable in person; when the activity is private, for example, in the home; or when it relates to feelings, for instance, emotions or motivation. For example, Jang et al. (2016) used diaries with interviews to collect data about users' experiences with smart TVs in the home as compared to within a controlled lab setting. The study in the home was conducted over several weeks during which participants were asked to keep a diary of their experiences and feelings. Surveys were also collected. This mixed-methods study informed the user experience design of future systems.

Diaries have several advantages: they do not take up much researcher time to collect data; they do not require special equipment or expertise; and they are suitable for long-term studies. In addition, templates, like those used in open-ended online questionnaires, can be created online to standardize the data entry format so that the data can be entered directly into a database for analysis. However, diary studies rely on participants being reliable and remembering to complete them at the assigned time and as instructed, so incentives may be needed, and the process has to be straightforward.

Determining how long to run a diary study can be tricky. If the study goes on for too long, participants may lose interest and need incentives to continue. In contrast, if the study is too short, important data may be missed. For example, in a study of children's experiences of a game, Elisa Mekler et al. (2014) used diaries to collect data after each gaming session in a series. After the first few sessions, all of the children in the study showed loss of motivation for the game. However, by the end of the study, those who completed the game were more motivated than those who did not complete the game. Had the data been collected only once, the researchers may not have observed the impact of game completion on the children's motivation.

Another problem is that the participants' memories of events may be exaggerated or detail is forgotten; for example, they may remember them as better or worse than they really were or as taking more or less time than they actually did take. One way of mitigating this problem is to collect other data in diaries (such as photographs including selfies, audio and video clips, and so on). Scott Carter and Jennifer Mankoff (2005) considered whether capturing events through pictures, audio, or artifacts related to the event affects the results of the diary study. They found that images resulted in more specific recall than other media, but audio was useful for capturing events when taking a photo was too awkward. Tangible artifacts, such as those shown in Figure 8.12, also encouraged discussion about wider beliefs and attitudes.

Figure 8.12 Some tangible objects collected by participants involved in a study about a jazz festival
Source: Carter and Mankoff (2005). Reproduced with permission of ACM Publications

The experience sampling method (ESM) is similar to a diary in that it relies on participants recording information about their everyday activities. However, it differs from more traditional diary studies because participants are prompted at random times via email, text message, or similar means to answer specific questions about their context, feelings, and actions (Hektner et al., 2006). These prompts have the benefit of encouraging immediate data capture. Niels van Berkel et al. (2017) provide a comprehensive survey of ESM and its evolution, tools, and uses across a wide range of studies.

Interaction Logs, Web Analytics, and Data Scraping

Interaction logging uses software to record users' activity in a log that can be examined later. A variety of actions may be recorded, such as key presses and mouse or other device movements, time spent searching a web page, time spent looking at help systems, and task flow