3 Discovering What to Build

Discovering what to build is often the hardest part of creating new products that seek to fundamentally change the ways in which people interact and communicate. By ensuring that a new concept is rooted in users' behaviors and needs, often actually designing and building the resulting system turns out to be quite simple. Getting the right concept, however, is often difficult and the result of much trial and error. The goal of good generative research is to increase the percentage of hits that get taken forward to successful solutions that can help people in their daily lives. There are many ways to be inspired, but we have found that engaging with people out in the real world and learning from their experiences is the best way to get that spark of an idea that leads to an invention.

Following the general pattern of user research outlined in the previous chapter requires some modifications for the mobile environment. Mobile interaction takes place in the contexts of daily life, often in quick bursts and in parallel with other tasks such as waiting for a train, navigating a city, or collaboratively planning activities with friends and family. We have found that observing these interactions between people, media, and context can often be the best inspiration for new concepts, as we can find the key tasks and behaviors that are currently unsupported by today's mobile applications and services. We then use insights from our observations to help in designing and building that new experience. This requires getting data from a variety of places and situations outside of the lab environment and engaging with users to understand their lives. It also means using creative methods to "observe" users' interactions at times when a research team cannot be looking over their shoulders.

Paul Dourish from UC Irvine (2006) saw the way that direct observations in the world can create ideas for new products and services and called these findings from research most valuable as "inspiration" for design in a much-discussed 2006 paper at the CHI conference on Human Computer Interaction. These inspirations differ from often dry and detail-free design guidelines that historically came from field research

in design. By directly sharing observations of users and quotes from interviews, researchers and designers can be inspired to build off of subtleties from field observations in creating new concepts. We wholeheartedly agree with this approach and view our early-stage ethnographically inspired work as a means to get smart in a design space and to get the creative design concept ideas flowing.

When starting to work in a new space (or even an old one after some time has passed and new practices have evolved), it is important to understand how people currently interact with technology and each other in the domain you are interested in. Understanding what does and does not work for them is often the best way to invent new concepts to help with unmet needs or to make sure your design ends up fitting into the ways in which people want to interact with each other and with the content provided by your system. Participating in this type of research not only prepares a team to design new concepts in that space but turns them into domain experts, or at least advanced novices, to be consulted on further projects in that space. This knowledge often stays relevant for some time after the studies are complete, as they are meant to create an understanding of deep human behaviors and interactions that are less dependent on the current state of technology. We have applied these tools in many domains over the years before creating designs and prototypes in these new areas.

In 2006, Frank and Crysta Metcalf became interested in location-based services and the concept of sharing one's location with close friends and family. However, we saw many problems with existing location awareness tools that shared absolute location at all times and seemed to amount more to tracking than facilitating healthy social interactions around locations. We hoped that we could create a system that would fit in better with people's true uses of location sharing in their current social interactions and would enable people to engage in new opportunistic interactions with the new information. We wanted to start by looking into the ways that people used location in their daily lives in order to provide some inspiration for the design of new concepts. By understanding the reasons for sharing location in everyday phone calls, we hoped to better understand how location sharing fit into patterns of everyday interaction out in the world. With this knowledge, we would be in a better position to invent new services that fit with the nuances of everyday life and were better aligned to people's real-world privacy concerns. We believed that this understanding would be critical in exploring why existing location-based services were failing to work in people's lives and would lead to insights to allow us to create new location sharing concepts that fit with people's motivations for sharing their location with others.

We devised a study in which we gave a set of diverse participants phones that would allow them to record their phone calls (with permission of all parties) for a week and

save them automatically as time-stamped files on a memory card in the phone (Bentley and Metcalf 2008). Midweek and at the end of the week, each participant would FedEx us the memory card from their phones so that we could transcribe the phone calls and interview them about their location sharing. We were interested in the ways in which they talked about location with each other, the purposes for this sharing, and their effects.

From this study, we learned that people in strong-tie social relationships often share the details of their current location and near future plans with each other over the phone and in person. This sharing allows close friends and family to have a fairly accurate mental model of where their friends and family are at most times of the day. However, what is often unknown is the transition time between these locations. Participants in our study frequently asked questions to confirm their exiting mental models such as "Have you left yet?" or "Are you home already?" We also saw that being unclear about someone's location sometimes led to no communication taking place at all, for fear of interrupting someone when they were at work or driving.

With this finding, we created a concept, Motion Presence, which helped people to know when their close friends and family were transitioning between places, without sharing the actual details of their locations with each other (Bentley and Metcalf 2007). In this concept, an augmented phone book showed if particular contacts were "at a place" or "moving" between places and for how many minutes they had been in that state. However, it did not share any details about what that place might be. In later field trials, we saw that participants were able to get an understanding of availability from the data provided and used this information to decide when to interrupt their friends and family with a phone call. They also appropriated this information for all sorts of unexpected uses including getting more time at their current activities if they saw that someone was running late or using the data to know that a loved one was safe and moving about as they usually do throughout a night shift or commute home late at night. The insights from this concept eventually led us to the presence-enabled phone book in MotoBLUR which is now used by millions of people to stay up to date with their friends and family and to coordinate activities.

Another idea came after mobile devices began to add significant amounts of user storage space in 2004. Devices with Secure Digital (SD) cards were on the horizon and we saw the inevitable future where phones and MP3 players combined into single devices. Because of this, we became quite interested in how people interacted with their music collections. Frank, along with Crysta Metcalf, Gunnar Harboe, and Vivek Thakkar at Motorola Labs developed a study to investigate how people selected music to play and how music was stored and acquired (Bentley, Metcalf, and Harboe 2006).

We knew that interacting with large lists of content on mobile devices was going to be difficult, as was already seen with the scroll wheel on Apple's first iPod. We knew that there had to be a better way to select the right music for a particular context and went out to study this phenomenon in daily life.

In the study, we visited people's homes and learned how they organized music, selected music for particular events, and how they acquired new music. We learned that music selection was often an iterative process whereby a user would pick something to play and then desire something a bit different—a little faster or a little slower, newer or older, something they had not heard in a while, or their latest acquisition. We also saw the desire for the ability to seek out "more like this" in particular ways: perhaps more from that artist, more in that genre, or more from that time period. We observed that this process of music selection follows a satisficing behavior where people would stop when they reached some music that generally fit into what they felt like listening to at a given time.

From these observations, we created two concepts. The first was the Metadata Knob, a dial that allowed a person to tune her music collection just like a radio (Bentley et al. 2008). Just set the knob to beats per minute and dial in the music for your current mood. As you turn the knob, you hear different music start to play, and you turn until you hear something that you like—just like a radio. Less than a week of development later, we had our first prototype. Our second concept was a play tree, where instead of a linear playlist that goes from one song to the next in a predefined order, users could see a list of next possible songs based on different metadata attributes of the currently playing song (Harboe et al. 2008). One option might be another song from the same album, one might be from another song in your collection that was released that same year, and another might be one with the same tempo. If the user did nothing, the next song in the list would be selected automatically. But if they wanted to be a bit more interactive, clicking on their choice for the next song was an easy way to help direct the music playing to be more like what they were looking for.

These examples are just a few of many from our work at Motorola. In our class at MIT, our students follow a shortened version of this process in the first week of the class to help seed their minds with data from real user behavior in the areas in which they are interested in developing applications. Some projects have investigated supermarket shopping, overseas travel planning, couch surfing, or alarm clocks. By getting a quick understanding of how people currently perform these activities, students are able to ground their design work in more than just their own experiences and create better overall systems. This process does not need to be long (we will discuss ways to shorten it at the end of the chapter), but it is valuable before designing or building

technology in a new space. You will often be surprised how your assumptions differ from reality and how much Lund's (1997) mantra—"you are not [like] the user" of your intended system—holds true.

The Process of Generative Research

How does one get from a space of interest (e.g., music, location sharing, intergenerational communication, etc.) to a list of potential solutions? The work of Crysta Metcalf and the entire team in our lab at Motorola have refined this process over the past decade. Together, we have tried many different methods for understanding current behaviors in an area and applying this knowledge to create design guidelines and ideas for new concepts. This process takes the researcher on a path from well-defined research questions to methods, data collection, analysis, and ideation. A study that is rigorous and publishable might take a few months of work, but discount methods allow for much faster cycles and can take as little as a day or two. Depending on the goals of the team, this stage in the process can take an amount of time anywhere in between those extremes.

Defining Research Questions

The first step in any research process is to formalize the questions that need to be answered. These are generally quite broad questions about your domain of interest and will lead your team to specific insights on user behavior. Because these studies tend to be small and qualitative, these questions should be more about understanding the breadth of use (the how and why) instead of the quantities of use (how much and for how long). Since the rest of the study and the data you collect will be based on these questions, it is important to give some thought to all aspects of your domain that you might find interesting. It is always easy to prune back the list of research questions when designing interviews or probes, but this first list should be as all-inclusive as you can get. Lindlof and Taylor (2002) also include a detailed discussion on defining research questions for qualitative communications research.

The use of mobile systems is often contextual, and our research questions have tended to focus on understanding these contexts of use. They often focus on where and when given interactions take place and both the trigger that started the interaction and any follow-up action that was taken because of specific interactions. We are also usually quite interested in the social interactions between people. Sometimes this involves phone calls or messages between people but can also include interactions

with people who are present, such as collaboratively selecting music or deciding on a restaurant in a group while mobile.

A few research questions from some of our studies can be used as a guide:

Intergenerational Communication Over Distance (Bentley, Harboe, and Kaushik 2010)

- How do participants communicate with remote family/closer relations? What tools do they use?
- What are the barriers to remote communication? What challenges do people face in maintaining or engaging in remote communication?
- What communication tensions and obligations exist surrounding remote communicate for the elderly?
- What are differences between communication at a distance and communication when remote relatives visit?
- What artifacts in the home serve to promote remembrance of and communication with distant family?

Use of Location in Phone Calls (Bentley and Metcalf 2008)

- During communications with others, what location and activity information is provided?
- Under what circumstances is this activity and location information disclosed?
- Why do people disclose location and activity information?
- How are disclosures of activity and location similar? Different?

Music Usage (Bentley, Metcalf, and Harboe 2006)

- What breakdowns exist in today's music experience in both independent and co-located situations?
- What contextual (from the past and the present) metadata can be used to address the breakdowns in today's music experience as identified in previous question?
- How can this contextual information be used to enhance today's music experience?

As you can see, these research questions are broad and aim at uncovering new insights about user behavior in the target domain. By keeping the questions broad, there is a greater chance that the research team will uncover unexpected findings that will lead to novel design ideas. These questions rely heavily on the context of the user and interactions in particular spaces and situations. With mobile services, integrating new experiences into the contexts of daily life is often the hardest part to get right.

From the research questions, a study must be devised that will answer the questions and lead to new design insights. There are many methods that can be appropriate

depending on the types of questions that have been identified. Many of these methods serve to uncover interactions that are not directly observable by research teams or to further understand particular contexts of use. In the next few sections, we will explore some of these methods with examples of how they have been used in work throughout the research community and in our class at MIT.

Using Probes

Mobile behavior is difficult to observe. The most interesting uses often occur at times or places that are not directly observable by a research team. Often much use is contextual or emotional and cannot be replicated in a lab or home interview. Understanding these contextual uses is critical to understanding how today's technologies are used and the situations that lead to joy or frustration. As Grimes and Harper (2008) discuss, designing from observations of joy and celebration can create new concepts that are fun to use and build off of the best of our interactions with the world. However, since mobile interactions are often short and particular details of a mobile interaction are often unmemorable, simply asking people to recall particular instances of a behavior in an interview will not yield the detailed (or even accurate) data that is needed for design. And it will rarely recreate the emotions of the instant that are often critical in interpreting an interaction. This is where probes come in.

In short, a probe is a piece of technology that is given to a user in order to help researchers observe parts of their life that are otherwise hidden from interviews or direct observation. The original cultural probes, as deployed by Gaver, Dunne, and Pacenti in 1999, consisted of a camera, postcards, and a map. Frequently, studies using these methods are called "Diary Studies" as participants often keep some sort of journal of their interactions and environments throughout the course of the study. The details that are logged more closely to an interaction are more likely to be accurate than a recollection some days or weeks after the event. In this way, researchers can get firsthand accounts of behavior as close to the time of interaction as possible. The overall goal of a probe is to be as simple as possible for the participant to record in situ, while still providing meaningful data to the research team.

We can classify probes into several buckets. First, there is a set of probes that are meant to simply serve as memory aides. These probes often take the form of a log or voicemail diary. Whenever a participant does a particular action, they are asked to log it. Usually in the log, they are asked to answer specific questions that are of interest to the research team, such as the time and context of the interaction as well as specific questions that follow from the overall research questions. For example, in a study on

intergenerational communication over a distance, we asked our participants to call into a voicemail right after any communication they had with their parents/children to tell us about the details of that interaction and what was shared while it was still fresh in their minds. Sometimes, these probes can be automated. Many mobile platforms now allow for monitoring of call state or messaging history, and automatic reminders with links to call into a voicemail or complete a survey can be presented immediately following the activity of interest to help increase response frequency while memories are still fresh. With these probes, we have found that keeping a regular frequency of calling (e.g., nightly) can help improve the likelihood that participants will provide data and not forget to call. Frequent calling also allows research team members to know when a participant has not provided information (even if all that they report is that they did not do anything you were interested in that day, the participant should still call in daily) and to remind the participant to keep calling in or completing surveys.

Another type of probe is meant to add visual information for the researchers that is not present in a verbal interview after the fact. These visual probes allow researchers to see into the lives and interactions that their participants experience. Often, these probes include photos or videos taken at the particular times and places of interaction. This can help researchers to better understand the context of an interaction and other forces that might be at play surrounding the user. For example, are there other people around that might have influenced the actions taken, are there other devices besides the phone involved in a particular interaction such as a paper phone book or computer, are there particular places that are frequently used for a type of interaction? Seeing the context can provide the types of insights usually gleaned from direct observation of users without the need to follow them day and night. These probes follow directly from the types of probes that Gaver's team used and help researchers to see into the lives of their participants and the sometimes unexpected contexts of use.

A final category of probes involves giving participants a piece of technology to see how they would react in a given situation. This technology is usually created to help answer a specific research question that might be impossible to answer without it. Consolvo and Walker (2003) at Intel Research applied the Experience Sampling Method (ESM) to mobile technology. In an ESM study, questions for a user pop up on a mobile device throughout the course of the day. In one study, the team at Intel asked users to indicate what fidelity of location (e.g., GPS coordinates, neighborhood name, city, etc.) they would share with a particular contact at the moment of interruption as a way to better understand privacy concerns around location sharing.

In a study in our lab at Motorola, Harboe and team (Harboe et al. 2006) placed a speakerphone between the homes of two friends who were watching the same television program in order to explore how people would converse over a distance while watching the same content. These types of probes are quite useful when studying phenomena that are not yet mainstream and in seeing how people would respond in these new situations. However, it is important to take the novelty effect into consideration and investigate how people continue to respond to these systems over repeated use. Insights from using a technology probe may be more accurate than an interview or focus group, since participants have actually used the concept; still, it is important to recognize that this use remains an initial impression and is not representative of use over time.

Logging and Conversation Analysis

Since mobile interactions are often quick and take place in parallel with other activities, the details of these interactions are often not memorable. In many cases, the data that researchers need to inspire new concepts relies on specific details of how people are interacting with their content and with each other. Data such as particular lines of conversation or specific steps taken in an application are quickly forgotten, even if participants are required to create a diary entry or voicemail directly after their interaction is complete. In these situations, automated logging or conversation analysis can help researchers to get at the missing data and complete the picture of the user's behavior.

Many mobile platforms allow for recording phone calls directly from a service running on the device or with a simple user interaction. When both parties have agreed that recording of phone calls is acceptable (policies vary by state and country), having access to this information can be a valuable added source of understanding for how users are planning, navigating, and sharing information with others. Alexandra Weilenmann of the Viktoria Institute pioneered these techniques back before audio recording on mobile phones was simple. With portable tape recorders, she had users record their conversations over several days and analyzed the data to understand how people talked about location with each other. She and Peter Leuchovius (2004) observed the use of personally meaningful location descriptors such as "where we met last time," which could be used in designing systems for micro-coordination: the iterative task of planning, and then actually meeting with other people. These types of in-group location descriptions could also be used to mitigate privacy concerns as third parties are unlikely to be able to interpret this type of location data. In another

study on availability, Weilenmann (2003) observed the use of location words in statements such as "I can't talk right now, I'm in a fitting room." This information was used by many researchers and mobile application designers to inspire various location sharing and mobile awareness systems that could help moderate availability, and it helped to tie together activity and location as frequently interchangeable descriptions of one's availability. These lines of conversation could likely not have been obtained in any other way, as study participants would probably only remember the vague details of their conversations if asked about them later. Here, the particular details in real-time speech provide the interesting data for analysis and new concept ideation.

In our own research on mobile location sharing mentioned at the start of the chapter, Frank and Crysta Metcalf (2007) at Motorola had participants record their phone calls with consenting friends and family and also analyzed location disclosure practices. We observed that people usually knew the basic mobility patterns of close friends and family, but often asked about the other party's location or shared their own to confirm transitions such as getting to work or being on their way to meet in person. This helped inspire many of our location concepts including Motion Presence (discussed in chapter 4) and the Contacts 3.0 service, which became the phone book in MotoBLUR in its commercial release.

Existing repositories of content can also be mined to understand user behavior. Researchers from corporations and academia have analyzed content on Flickr, YouTube, and Twitter to understand how people share content and build relationships online. Oskar Julhin, Arvid Engström, and Erika Reponen (2010) from the MobileLife Centre in Stockholm and Nokia Research analyzed the content of videos uploaded to mobile video sharing sites such as Bambuser and Qik in order to discover the types of media that are shared in these systems. They observed the relatively poor production quality of the videos, including transitions and poorly produced beginnings and endings. Their design recommendations include ways for mobile video broadcasting systems to improve these aspects of production and to tailor interaction with these systems to the types of content that are created while mobile, which they observed to be different from video content created while seated at a computer. The resulting systems would lead not only to better content on the site, but to users who are now better educated in the art of film production.

When building your own systems, building in the appropriate instrumentation so that you and others can analyze use or perform content analysis is important so that everyone can benefit from understanding how technology is being used in the world. The Facebook Data Team (2010) often releases data derived from their own content analysis, such as a recent exploration of voting practices in the 2010 midterm elections

in the United States. For example, they were able to predict sixteen of the twenty close races just based on counting Facebook "likes" for various candidates. Similar content analysis can be performed on many existing systems. More information on instrumentation can be found in chapter 9.

Logging interactions and communication can help researchers to better understand current behavior and use these findings to inspire new solutions that are tied to real users' lives. However, this raw data can be open to many interpretations. All that can really be learned from a log file is that a particular phenomenon occurred. No explanation of this action can reliably be given without more qualitative data. Explanations of these practices can develop from combining the log data with follow-up user interviews, leading to deeper insight that goes beyond the raw data and provides useful insights into why certain patterns of use occurred.

Home Tours/Field Visits

A person's home says a lot about who they are and what they consider important. The organization of items in the home can also say a great deal about how participants use various types of media. Data from a home tour can serve to put data from other methods in context, showing the places where interaction occurs as well as the objects and information that are present in that environment. "Home" tours can also include other environments of interaction such as workplaces or transit routes to better understand the environments in which users interact with their devices and with other people or content. When studying topic areas as diverse as photo use, music use, television, or intergenerational communication, we have employed the home tour as a part of our study design to learn more about how these aspects fit into our participants' lives. Students in our class have performed field visits in grocery stores, dorm rooms, coffee shops, and many other locations around Boston to better understand the context of various activities from grocery shopping to travel planning and mobile game playing.

In our intergenerational study exploring practices of communication across distance and generation, we started with a tour of each participant's home so that we could understand the variety of places and devices that they used to communicate (Bentley, Harboe, and Kaushik 2010). As we went through their homes, we asked about the last times that they communicated in each room or on a given device. We were easily able to learn about the places from which they liked to communicate and the factors that led to them choosing a particular communication device or location. We were able to observe how some locations created spaces for quiet, comfortable

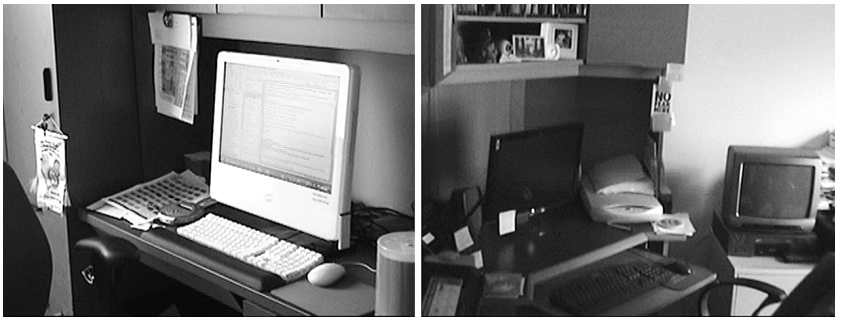

Figure 3.1
Photos from the home tours in the intergenerational communication study. Through our home tours, we discovered the importance of place in communication and the reduced use of communications technologies that resided in out of the way places such as extra bedrooms or basements. This led us to thinking more about the importance of place in communication and ultimately, in combinations with other findings, to the concept for Serendipitous Family Stories.

communication while others led to frustration or distraction. After the initial interview, participants called into a voicemail nightly to discuss their communications. Because we had visited their homes, we could visualize the places from which they reported making phone calls and the artifacts in their homes that led them to think about their relatives at a distance. Without the home tour, much of this information would not have been understandable from a static voicemail. We were also able to learn the diverse ways in which our participants integrated communication into their lives. When asked to show us where he communicates, one participant asked us, "You mean where the phone hangs on the wall?," while another participant talked about the wide variety of places both in and out of the home she communicates from including trains and tanning beds. Understanding these perspectives and visiting the environments where communication occurs helped us to better understand how place plays a role in communication, especially for older adults, and led us to focus on place-based content creation in our resulting Serendipitous Family Stories system (Bentley, Basapur, and Chowdhury 2011). In this system, users could leave video messages for friends and family in particular real-world locations. When recipients came near one of these locations, they would be automatically notified that a story existed there and would be able to view the video. This became a way to share family history in the rich context of the world in which it occurred.

In our music study from 2004, we also utilized home tours as a method (Bentley, Metcalf, and Harboe 2006). We were interested in the places in the home and car where

Figure 3.2
Photos from home tours in the music study. We observed the piles of physical music that our participants kept in various locations in their homes, cars, and gym bags, thus creating music zones where only particular music was accessible (left). We also observed the importance of gifted music in its original gifted packaging (center) and the ways in which our participants prominently displayed their favorite music in the home (right).

participants stored music as well as how they selected and played music in particular contexts. Most participants in this study still had fairly large CD collections as digital music and the Apple iPod had not reached a large market by this time. We could see how the music at hand led to particular music choices in certain contexts and how favorite or recently acquired music was kept at hand and regularly moved between car and home. We saw participants who got stuck in music listening ruts in their cars because of the limited set of CDs that were there as well as their strategies for getting out of these ruts and into new music or old favorites by rotating their collections. All of this information helped us to understand the physicality of music and some of the positive aspects of that physical nature that might be exploited in digital music systems. It led to particular insights in maintaining easy access to recently purchased music and to inventions like the Play Tree or Metadata Knob that could let users branch out from their currently playing music into other parts of their collections.

When conducting a home tour, it is important to focus on areas that directly answer your research questions. Often, these questions center on the places where a particular type of interaction occurs. What distractions might be present in that environment? Why is one area of the home used for this activity over others? What other items are placed near your items of interest? Through careful observation and non-leading questions that speak to the research objectives, researchers can understand how the environment impacts communication and content consumption behaviors.

These types of tours are often best conducted during an initial interview before the official data collection takes place. By seeing the home, the research team can build

a deeper understanding of their participants' lives. This understanding helps to place in context statements made by participants during the later data collection phase of the study.

Home tours are also best conducted with a pair of researchers. We often have one researcher videotape the tour while another takes the lead in asking questions and following up on statements made by the participant. We often alternate the interviewer and videographer roles between researchers and the videographer is always free to follow up on a question or ask something that the primary interviewer may have overlooked. Depending on the depth of the interview, home tours usually last about half an hour and are an easy way to quickly learn about a participant's environment.

Task Analysis

While many mobile interactions occur naturally in the world, at times it is useful to create a contrived situation in order to observe how someone interacts with a particular piece of technology. To do this, researches often perform a task analysis. Beyer and Holtzblatt (1998) explored this topic in their paper and book on contextual design. They commented on the power of watching people perform tasks in the contexts of daily use to help researchers better understand their work processes and how information and materials flow while accomplishing a task. We have applied methods derived from this work in attempting to understand people's interactions with music and existing mobile applications.

In our study on music use at Motorola, in addition to home tours and interviews, we asked participants to perform some directed tasks for us (Bentley, Metcalf, and Harboe 2006). In an initial interview we had asked them about recent episodes of music search and the scenarios in which they have looked for particular music in the past. Then later on, while in their homes, we asked them to act as if they were in similar situations and to select and play music for these particular scenarios. For example, one user was asked to pick music to play while studying. Another was asked to pick music to play for a card night with friends. These were tasks that they reported normally undertaking at home, so we were able to see how they used the music piles that they had around them and the devices available for playback to select and play what they thought was the appropriate music for each particular scenario.

During a task analysis, it is common to have participants "think aloud" and verbalize what is going through their head at any given time (as described in Lewis 1982). This helps researchers to understand their thought process and why they are choosing to do the task in a particular way. The think aloud method is another way to under-

stand what a user is thinking at various points of an interaction and not just what they are doing. By thinking aloud in real time, users are more likely to be stating what is actually going through their minds, rather than making up explanations if they are asked later why they acted in a particular way. Often, it is a bit uncomfortable for users to constantly verbalize what is in their head. Proper warm-up exercises, like having participants think aloud while trying to reload a stapler or setting an alarm clock, can help in warming up participants to speak while performing a task. Often, researchers will have to remind the participant to continue to speak while performing tasks. It is important to hear participants' thought processes as they go through each step in order to uncover usability issues and places where existing systems do not meet the intent of the user, such as when our music study participants wanted music that was a little slower or faster than what they were currently playing.

Task analysis can also be conducted out of context. In final interviews, we sometimes ask participants to show us how they have been using a particular system to help us clarify questions we might have from a voicemail log or interview. Seeing a participant interact in front of you, even if it is a task that she would normally undertake while mobile, can help you to understand the details of an interaction that she might not otherwise discuss in an interview or log.

From this analysis a number of models can be made. Beyer and Holtzblatt (1998) introduce several in their book *Contextual Design* that are quite useful, including diagrams of how information or content flows between users or places. These diagrams can be used to better understand how particular people or places encourage specific actions. For example, in our study of music use in the home, we saw how music tended to flow to specific places and "stick" there—for example, CDs brought into a car or bedroom. These findings helped us in brainstorming ways to get people out of their music "ruts" and on to different collections of music when they wanted something new.

Semi-structured Interviews

While other methods help to uncover details of use or interactions within a particular context, interviews still serve as a primary means to make sense of data that has been collected and to learn about how data collected during the period of the study compares to more typical patterns of behavior. All of our studies have included some type of semi-structured interview to better understand our participants' behavior.

Most of our interviews tend to be semi-structured so that we can explore areas of interest that our participants bring up in response to our discussion. These methods are derived from years of interview protocol design in fields such as anthropology.

The references for Bernard (2002), Angrosino (2002), and Ervin (2000) can help in conducting interviews and properly formulating non-leading questions.

We tend to focus our interviews on particular instances of interaction in our domains of interest (music use, phone calls with a particular person, etc.), usually starting with the last time our participants have interacted in a way in which we are interested. We then ask follow-up questions to better understand the context of that interaction and what happened next. As demonstrated in years of anthropological research, asking about the last few interactions can better help us to understand frequency of use and to have something to check against the data we collect during the rest of the study. This allows us to see if the data we collect during the study is typical of interactions that were performed before it without relying on people's poor abilities to recollect frequency data.

In final interviews, we often ask users to clarify data that we collected with other means, such as through a diary study or through logs from a mobile phone that they have been using. We use the existing data as a probe to help users remember the given context of interaction. The more data we have logged about the interaction, the easier it is to help users remember that particular instance. Reminding them of the day, day of week, time, and people involved in the interaction can help. Where additional data, such as what was discussed, is present, these details can also help jog a participant's memory. Showing participants a photo that they took and asking for more details around why, when, and where it was taken will lead to much more detailed and accurate results than simply asking participants to remember "the last time they took a photo." Rich visual cues from captured media and other log data can be the bridge to a memory that will result in a successful interview and add key missing explicatory data to the analysis.

Recruiting Users

For most generative research, we try to recruit between seven and twelve participants for each study. We have found that this number allows us to see a wide variety of use, but when using more than this number we begin to see a majority of repeated data with extra participants, which does not broaden our understanding of practices in the domain of interest. In order to achieve the broadest data set possible, we actively recruit participants that are as different from each other as possible. This generally means different ages, occupations, income brackets, genders, ethnic backgrounds, and technology use. For studies that examine a particular topic, we try to recruit people who have a wide range of use for that particular topic. For example, for the study in

intergenerational distance communication, we recruited some participants who spoke multiple times a day and other participants who emailed or talked only once every few months. Finding patterns across participants who are very different often points to more fundamental human needs that will lead to more universal application and service ideas that can work in many peoples' lives.

Recruiting is one of the most important steps in executing a generative research project. The data that is collected and the findings are only as good as the participants that are selected for the study. In the end, the data is only valid for the participants you interview and observe. If all participants are male computer science college students who are eighteen to twenty-one, it is hard to generalize the results or uncover a wide range of behaviors.

It is often difficult to find this perfect mix of participants on your own. Emails to extended networks, posts on mailing lists or Facebook pages, and putting up posters at grocery stores or college campuses can only go so far in attracting a representative group of users. Often, we need to use a professional recruiting agency to find our participants, especially when there are geographical or phone-related constraints (e.g., participants with Android phones living in Chicago with older relatives in Florida). While a recruiting firm can cost more money than trying to recruit on your own, the resulting diverse participants that they can find are almost always worth the expense in terms of the richness of data that can be collected.

Generative research is about being inspired to create new concepts. Therefore, having the broadest range of participants will give your research team that broadest set of behaviors to observe and use for design inspiration.

Conducting the Research

Our studies are generally structured with an initial interview at the beginning of the study to get to know participants and build rapport. Often the middle part of a study will involve gathering data from participants through logs and voicemail diaries, and studies will conclude with interviews to clarify the details of data collected throughout the study and to ask any final questions. Most of our studies run for about two to four weeks so that we can observe a variety of behaviors that occur and begin to understand interactions that are frequent, while still having the opportunity to learn about more rarely occurring but impactful events in our participants' lives.

We generally take two to three researchers into the field for any given field study. There are a few reasons for this. First, we would like as much of the team as possible to get to know the participants in the study and we would like the participants to get to

know as many of us as possible. Grounding researchers in at least some details of each participant creates a familiarity that is useful during analysis. It also helps us build rapport with the participant and introduces them to anyone who might call during the study to remind them to file a voicemail probe or schedule a final interview.

Second, having two researchers allows for more of a dialogue with the participant and allows for someone who is not focused on coverage of the primary questions to focus on additional follow-up questions based on the participant's responses. This can take some pressure off of the primary interviewer and almost always leads to more complete data.

Finally, we generally audio- and videotape all of our interviews so that we do not have to frantically scribble notes during the interview itself. The second or third researcher can handle the technology and make sure everything is captured. This is especially helpful during interviews that contain home tours as the visual record is critical to the analysis of this portion of the research.

We also strongly believe in conducting interviews in person. While this may involve extra time and travel expense, meeting with someone to establish a relationship and seeing his environments, media, and devices provide countless benefits in interpreting the data that is collected in the context of his life. We also often ask users to bring out specific devices or media that they mention in an interview, and this would be more difficult over a distance. Some promising work has been conducted using Skype video for conducting research remotely, but this requires users with a bit more tech savvy and still does not afford the ability to look around a participant's environment for additional cues.

We generally start a new research project with one to two pilot participants who are usually not part of the reported data. This allows us to test our research protocol, add additional questions, check the length of interviews so that we can report approximate time commitments to participants and recruiters, and ensure that our research questions are being answered by the data that is collected. After a pilot study, we will often rework parts of the study in preparing for the actual set of participants. Small changes can still be made once a study is in the field, and we generally debrief as a team after each interview in order to update questions and discuss follow-up questions that we should add to the protocol based on the initial interviews.

Affinity Analysis

Often, a generative study can create thousands of individual pieces of data. That data could be a sentence or two that a participant says that pertains to a particular research question, a photo, an observation made by the research team, or any other informa-

tion that the team has collected that pertains to the questions that are to be answered. We prefer to conduct our analysis directly from the raw data instead of using researcher summaries or insights as a base. This gives us the ability to trace specific higher level findings back to specific users and statements and to be more confident that the findings result from actual actions that occurred and not from biased researcher notes that might be more selective in nature or more likely to interpret specific actions and quotes from users in particular ways. With all of the raw data in the analysis, the traceability back to the individual participant is always present should anyone question (or just want to better understand) a higher-level finding.

The process of getting from collected data, often in the form of audio or video files of interviews and voicemails, to data items for analysis involves some effort. We first listen to each recording and transcribe any bit that pertains to one of our research questions. Each relevant statement from a participant gets its own item for analysis, often just a sentence or two long. On the item, we note the participant that the data came from and in what context (Initial Interview, Voicemail from day 2, etc.). We also often note the timecode in the file where that statement occurred in case we want to pull out video or audio for later presentations. Each item is also numbered for later identification and printed onto a post-it note for analysis as shown in figure 3.3.

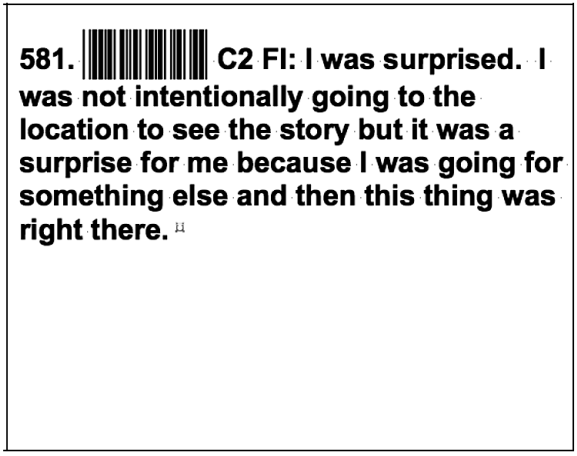

Figure 3.3
Example note from an affinity analysis on the Serendipitous Family Stories system. The note has a unique number, a bar code representing that number for easy scanning, the participant ID (C2), and the part of the study from which the data was collected (FI = Final Interview). This is followed by the direct quote from the participant. Hundreds (or thousands) of notes like this are then combined into an affinity analysis as shown in the following figure.

Once we have all of the raw data identified, we perform a grounded-theory affinity. The affinity is a bottom-up approach for iteratively making sense of raw data and developing theories of interaction based in the raw data. This analysis comes from the combination of methods from two different backgrounds. Jiro Kawakita, a Japanese anthropologist, created the KJ method of analyzing data (Scupin 1997). Seeing that standard ways of analyzing data in terms of existing theory and hypotheses were not working, he saw this method as a way to inductively build up meaning by finding interrelationships between items of data.

This affinity analysis method was popularized in the Human Computer Interaction community by Beyer and Holtzblatt (1998) in their paper and later book on contextual design. They adapted the KJ method and shortened the process by performing analysis on researcher observations instead of direct data from participants. We prefer to use the raw data for analysis so as to better control for researcher bias and to base findings in the actual quotes and data from our participants.

In an affinity analysis, the first step is to group data based on similarities that are observed. These first level groupings seek to describe the data and patterns that emerge across users who are saying or doing similar things. These similarities might be anything: two users trying to perform a similar action, examples of how place and other contexts shape an interaction, similar concerns or praises that participants raise, etc. These groups can represent any relationship within the data that the research team thinks is interesting and often evolve many times as the larger data set is analyzed. Often during the analysis a group that started small will grow as more supporting data is found. These groups are then broken down as nuances between the data items are more apparent.

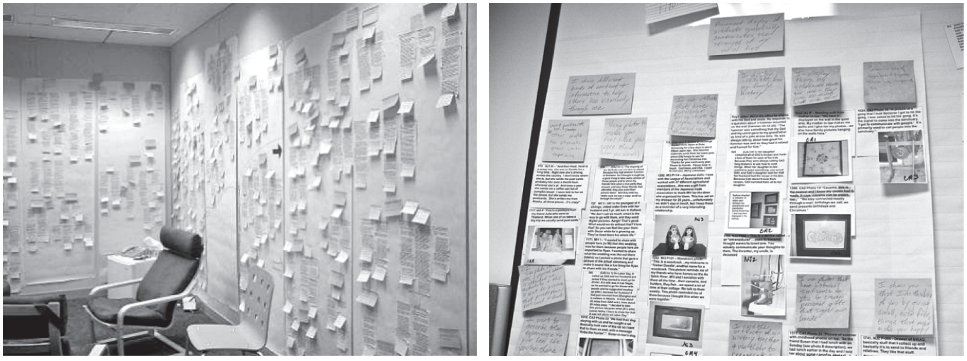

Figure 3.4
Photos from an affinity analysis showing the hierarchy that is developed as the data is analyzed.

In the end, we try to get the first level groupings down to piles of three to six notes from a variety of participants. At this point, we create an explanatory label on a new, differently colored note. This label serves to state the main message of the group, often in the first person. Examples of labels from a study on communication over a distance (Bentley, Harboe, and Kaushik 2010) include: "Gifts from my kids around the house remind me of them"; "Family hand-me-downs around my home remind me of my parents"; and "When I wear jewelry from my parents, I feel that they are with me." Each of these labels had between three and six items supporting them, generally from multiple study participants. When a variety of diverse participants start to say the same things or do similar actions, it becomes more likely that we have found something more generally applicable to a larger population instead of just the idiosyncratic behavior of one or two participants. In some studies we have seen seventy-year-olds and twenty-year-olds using communications technology in similar ways which helps us know that we are on to something much more generally useful.

The second level of grouping in the affinity hierarchy seeks to explain the data. Groups from the first level are combined together into sets that tell a larger story about the behavior that was observed. In the case of the three groups mentioned above, this could be a statement such as "Physical objects can create a sense of togetherness over distance." These larger groups help the research team to look beyond the individual notes and on to patterns of use that occur at a broader level. These groups can be seen as sub-findings of the larger research: findings that have some amount of support behind them from several clusters of lower-level notes. In a study with 1,000 data items, we can typically have twenty-five to thirty of these groups.

Finally, a third level is created that serves to tell a larger story of use. Usually a given study will have a small number of these high-level themes that can serve as categories to summarize the major findings in the study and as themes to work off of for inventing new concepts and validating existing ones based on the data that has been collected. Some examples of these high-level themes in our intergenerational communication study include "Sharing stories about daily life creates togetherness over a distance" and "The concept of the family is an important part of family communication." These themes helped us see the larger goals of our participants in maintaining their relationships across a distance and what they did to keep these connections strong when they could not see each other in person as often as they would like.

Once all three levels of the affinity are complete, we hold brainstorming sessions to invent new concepts that are based in the data. We place short descriptions of each idea on brightly colored post-it notes right next to the raw data or insight that inspired

them. From a study with 1,000 data items, we often create in excess of 100 design ideas. These ideas are directly tied to the themes and participant quotes in the affinity diagram. Examples such as the Metadata Knob came from observing a user trying to find music that was a bit faster than what he was playing and another user who was searching for something a bit newer than the currently playing song. The idea for Serendipitous Family Stories/StoryPlace.me came from a similar observation. One of our participants was walking past a theater and was talking with her mom about the importance of that place in their family history. We saw the opportunity to create these types of interactions serendipitously with location-based videos and notifications when users approached locations that contained videos that were shared with them. The ultimate concept used each of the high-level themes from the analysis to create a system that reinforced the memories and roles of the family, created communication that fit into everyday life, and created a feeling of being together with family members by placing video into real-world settings. While there are certainly other ways to arrive at concept ideas like these, we have found that the affinity analysis and using actual experiences to ground new concepts gives us the additional confidence that a new idea is useful for a broader population than ourselves. Also, the sum total of the data in the affinity and the resulting design guidelines help shape our new concepts far beyond the single note or group that inspired them. This allows the resulting solutions to more accurately fit into people's everyday lives and interactions with the world.

While sometimes we have one or two ideas that everyone agrees are absolute winners, often a prioritization process is necessary in order to identify the ideas that are the most valuable to take forward. These prioritizations can contain columns for applicability to current businesses and products, novelty/patentability, difficulty in building the concept, potential market size, and other factors. From this process, we generally have two to three high-quality ideas that we wish to take forward into functional prototypes. We have also found that the design guidelines generated can be quite helpful to other product teams developing concepts in related spaces (Metcalf 2011).

The next step for the highest-rated design ideas is often a quick prototype to validate the design hypotheses as well as some technical aspects of how they could be built. That will be the focus of the next chapter, but first we will focus on some ways to shorten the process of generative research when time does not allow for a more formal study.

Discount Methods

Often, research or product teams do not have time to complete a full, rigorous study but still would like to base their designs in user observations. In our class at MIT,

students have a week to run a quick study and only an hour or two in class to analyze the data that they collect. This does not allow for time to create a formal interview protocol, gather and transcribe hours of interviews, and create 1,000+ note affinities. But we have found ways to make the short time that they have worthwhile so that it improves their overall designs and grounds their systems in real-world behaviors and needs.

The design firm IDEO (Kelley, Littman, and Peters 2001) popularized the concept of a "deep dive," a quick way to put designers in contexts of real use, both by observing people and figuring out firsthand where existing products and environments failed their users. Their video from ABC's Nightline (Deep Dive 1999) in which they design a shopping cart is required viewing in many design schools and HCI classes around the world. Getting out in the world with a quick observation is a great way to ensure that new concepts are grounded in everyday realities. We encourage our students to take a bit more disciplined approach that lies somewhere between the types of observations shown in this video and the full, more ethnographically inspired, rigorous process mentioned above. First, we start with a list of research questions that get to the core of what the team wants to discover. From there, we ask them to spend an hour with people from their target audience watching what they do and then interviewing them to better understand their behavior. They can also ask participants to log behaviors or take photos in a log for five days. The data collected (often in the form of observation and interview notes) is then written down on post-it notes. From this hour of observations, students frequently have between fifty and one hundred notes. We then perform a mini-affinity in class where they group the notes to find patterns and explanations of use.

While a one-hour observation is not going to be as exhaustive as a full study with a dozen or more participants for several weeks, it tends to be a good start and gets students thinking about their target domains in new ways through the eyes of the people that they observed. Often, this leads to follow-up observations or the desire to observe more people in different contexts of use. In almost all cases, it helps students to see how some of their initial assumptions might not have been correct and that the range of behaviors in their domain of interest is different from what they imagined. This helps them to create stronger concepts that will work for a broader range of people, which is ultimately the goal of generative research. Some of our students in the past few years have explored a variety of routine domestic activities. In each case, by observing and talking with just a few people over an hour, teams were able to identify a range of behaviors much broader than what they had originally expected. In each case, ideas from these observations would become key differentiating components

of their final solutions. Often, students find different types of user segments that they had not expected or interesting ways in which people are already trying to accomplish a task. These insights dramatically improve the quality of their final systems and create solutions that are likely to work with a wider population than what they would have created without this generative data. Also, as in the shopping example discussed earlier, students often discover a key part of an activity that makes it fun and can turn that insight into a focus of their ultimate design.

Other methods, such as those that IDEO has outlined on their Method Cards (2002), can help to quickly explore current practices in a variety of scenarios and lead to quick observations or interviews that can ground a design in real-world behavior. Most of these methods can be completed in a few hours and can give design and research teams quick new perspectives on their domain of interest.

However, researchers must take care to understand that the data they collect in any of these rapid methods is just a small sample of the data that it is possible to collect. If only a few people are observed, it is easy for research and design teams to overly constrain their resulting design for the types of use that were observed and thus leave out many potentially common use cases. This can also happen with the use of personas or user-types derived from small amounts of data. Designers often focus on designing for those particular personas and forget that other types of users could exist that are hybrids of the personas or that represent entirely different types of use. These discount methods should be carefully used to broaden design thinking and open up new possibilities instead of narrowing resulting designs to conform to the small amount of data that is collected.