Studying and Designing Technology for Domestic Life

Lessons from Home

MK
MORGAN KAUFMANN

Tejinder K. Judge
Carman Neustaedter

CHAPTER

In-Home Deployments

9

A.J. Brush*, Brian Meyers*, James Scott[§]

Microsoft Research, Redmond, WA, USA; Microsoft Research, Cambridge, UK[§]*

INTRODUCTION

In-home deployment is a critically important method for studying technologies for domestic life, as it is only when these technologies are inserted into the reality of daily home life that many of their properties become fully apparent. An in-home deployment study can also be referred to as a field study, or *in-situ* study, where the “field” is someone’s home. In our view, the main benefit of an in-home deployment compared to other commonly used evaluation methods, such as laboratory studies or interviews, is *realism*. A deployment study takes place where people actually live, and the data that is gathered is from their normal lives. This chapter focuses on studies where each home in a study is an independent deployment site, while Chapter 10 provides more details on field trials with multiple connected homes (e.g., children and grandparents sharing an always-on video connection).

The two most common types of in-home deployments are studies of current behavior and prototype deployments. Current behavior studies seek to understand what household members currently do, for example, how often they are home together or how much energy they use. This may require deployments of non-interactive devices, such as for logging purposes. On the other hand, the goal of a prototype deployment is to study people’s reactions to and use of a prototype, as well as how the prototype influences their behavior. While current behavior studies sometimes lack some of the challenges of prototype deployments—for example, the occupants may not be required to interact with a new technology—they have similar complexity in the time and effort of deploying to homes and ensuring that data collected is valid.

This chapter will help in the planning and successful execution of in-home deployment by sharing insights drawn from our combined experiences running many different studies in the home and by using a recent, fairly complicated study, the PreHeat prototype deployment (Scott et al. 2011), as a concrete example. PreHeat is a prototype thermostat that automatically controls home heating using occupancy sensing and prediction. We deployed and studied the system in five homes over the course of several months. Prototype system development and debugging, planning the study, and running the study took approximately one year. During this time, we faced challenges that included accurately sensing home and room

Studying and Designing Technology for Domestic Life. http://dx.doi.org/10.1016/B978-0-12-800555-2.00009-5

occupancy, building a robust heating system that failed infrequently and recovered quickly if it did fail, and designing the study to fairly compare multiple long-running conditions. In this chapter, we reflect on those issues and provide take-home lessons from our experience.

IN-HOME DEPLOYMENTS

First, we discuss in general when in-home deployments are appropriate, how to anticipate the effort required when doing a study in the home, and how to deal with the challenges of in-home deployments. We provide guidelines and strategies relevant to all types of in-home deployments. We then focus more closely on challenges presented by our deployment of the PreHeat prototype and what lessons we gathered from that experience.

CAREFULLY CONSIDER THE VALUE OF AN IN-HOME DEPLOYMENT

Deployment studies are appropriate for observing current behavior over time or needing to place novel technology into a home environment to see how participants interact with it during their normal lives. The trade-off is that there is much less control over what people do during the in-home deployment study.

While in-home deployments result in rich data and can lead to very robust findings, caution should be used when deciding to pursue one. In-home deployments are challenging for a number of reasons. First, conducting them is time consuming. Typically, a home must be visited multiple times, including at the beginning of the study, end of the study, and sometimes during the middle of the study (e.g., if a prototype needs to be repaired). Second, the technology deployed must be robust enough to survive both the intended use and any unintended uses that will almost certainly arise during the study period. Lastly, there is little control over how participants use the technology during an in-home deployment. One might well find that one's technology is used more for some unrelated purpose rather than the topic of the study, or not used at all.

With these challenges in mind, we encourage researchers considering an in-home deployment to reflect carefully whether another method might be more appropriate or have a better cost/benefit trade-off. For example, laboratory studies are more appropriate for any research questions in which people must perform specific tasks, and a single in-home interview at several different households may provide enough data to understand current behavior (e.g., Brush and Inkpen 2007). More information on a wide range of study methods can be found in HCI textbooks, including Interaction Design (Rogers, Sharp, and Preece 2011) and Human-Computer Interaction (Dix et al. 2004). Information on field studies in particular can be found in the "Ubiquitous Computing Field Studies" chapter in Ubiquitous Computing Fundamentals (Brush 2010).

That said, we reiterate that in-home studies offer unparalleled *realism* and insight into the use of technology in real contexts. Therefore, while we recommend careful consideration of the challenges that accompany in-home deployments, there are often studies that just cannot be performed any other way with the same fidelity. In many cases, a full in-home deployment may be most appropriate only after conducting a number of other pilot investigations, such as surveys or lab studies, which help to refine the system to be deployed, the study methodology, and the hypotheses or areas of interest to be explored during the study.

LESSON

Look carefully at the full costs involved in an in-home deployment, and consider using other study methodologies (e.g., lab studies, logging studies, surveys), either as alternatives or precursors to an in-home deployment.

TIMEFRAME FOR AN IN-HOME DEPLOYMENT

In-home deployments typically last anywhere from a couple of weeks to several months, depending on the level of interaction the household members have with the deployed system. For example, a study of current behavior by silently logging energy data is invisible to residents and can likely go on for as long as one desires. At the opposite extreme, a study that requires participants to track what they ate or what time they went to bed might experience issues with participant fatigue. Often, instead of thinking in terms of days, it is more useful to consider the number of data points that can be collected. A study of a new alarm clock might only be able to collect a single data point a day, and thus, in order to collect enough data to have confidence in the evaluation, one might need multiple months of use.

Hnat et al. uses the analogy of a battery to think about how participant effort required during a deployment study interacts with the length of the study (Hnat et al. 2011). Essentially, participants have a certain amount of "energy" available to participate in a study. That energy can be spent in a short amount of time by asking them for a high-level of participation or spent over a longer period by requiring less effort from the participants. Gratuities, or payments to the participants, can also be used to adjust the amount of effort participants are willing to expend in a study. For example, if participants must self-report, they can be paid for each self-report or given a reward for reaching a minimum level of reports. However, if the study aims to report on the naturalistic usage levels for a prototype system, then a different gratuity solution is needed to avoid having the payment bias the results.

Finally, it is important to note that a research project will last much longer than the length of time the in-home deployment study is conducted. This is not only because of the preparation time, but also because it is valuable to pilot a study before deploying it into participant homes. Every study we conduct has an initial pilot in our own homes and often one or two friends or family members. The pilot study does not have to last as long as the real in-home deployment, but the researcher should pilot every part of the study, especially if he or she has different conditions. A pilot also helps to verify that the researcher knows exactly what will happen at each visit to the home (make a script and a checklist) and that all the necessary equipment and tools are acquired to successfully deploy a prototype or conduct a study of current behavior. Once the researcher deploys into someone's home, it can be difficult to schedule time to go back, so do everything possible to get it right the first time. We recommend always having extra extension cords, surge protectors, a small toolkit, and a camera. We photograph the space both before and after our deployment, even if we do not intend to use the photos for anything more than reminding us about each setup.

LESSON

Plan timelines for in-home studies that include time for study overhead, such as pilot deployments, setup and debugging time, deployment overheads, and teardown time. Bear in mind "participant fatigue" in the level of interaction demanded of participants during the study.

DATA PRIVACY CONCERNS

The data collected during an in-home deployment will vary depending on the research question. One important general consideration is how to handle privacy concerns. Think carefully ahead of time about how to analyze the data, and test this during the pilot phase. Ensure that all the data needed is being gathered and at the appropriate granularity for the analysis required. However, it is also important to avoid doing too much data logging, since the additional data can make it harder to focus on the data needed, and with additional data comes increased privacy concerns.

Because in-home deployments collect data from people as they live their lives, the deployments can often capture data in sensitive spaces. Sometimes these concerns are very clear, for example, audio recording in a bedroom (Kay et al. 2012) or always-on cameras in household public spaces, such as kitchens (Pousman et al. 2008). In other studies, the sensitivity can be less obvious. In the PreHeat deployment, we collected occupancy information that could reveal times when the home is unattended, wake/sleep times, when and for how long bathrooms are used, and so on. It is very important to have a plan around what data will be collected, who will have access to it, how it will be anonymized, and how long to keep it. This plan should be shared with the participant households so they fully understand what is being collecting and how it will be used. In most settings, an institutional review board (IRB) will require the researcher to address these questions.

While collecting data in a privacy-sensitive manner may take some creative thinking, there are many options available. One option we have used successfully was to offer participants an additional gratuity if they opted in to that aspect of the study we considered sensitive. In our Speech@Home study (Brush et al. 2011) we wanted to collect five hours of audio data to understand typical noise level in kitchens. We put a "record now" button on our prototype and offered participants an additional gratuity item if they recorded the extra audio data. Other studies have offered the option to review and delete data locally before it goes to the research team. The Lullaby sleep-capture system deployed on bedside tables used multiple cameras and a microphone. It offered three types of privacy controls, the ability to turn off recording, selective deletion of recorded data, and the ability to delete the last hour of recorded data (Kay et al. 2012). While such controls may skew the data captured, such a skew may be unavoidable in order to address privacy concerns.

LESSON

Gathering specific data in participants' homes is difficult without also unintentionally capturing additional data through accidental logging or by inference from intentionally gathered data. Take a strong and holistic privacy approach to protect private data, for example, by allowing participants to vet their data.

COST TRADEOFFS

When conducting deployment-based research, there are many tradeoffs to make involving the cost of the deployment. These costs are not just the monetary costs of the equipment or gratuities, but also include the costs in preparation time before the deployment, in management time during the deployment, and the risk of things going wrong that might necessitate re-running deployments.

When purchasing equipment for home deployments of either type (studies of current behavior or prototype deployments), we recommend significantly over-specifying the equipment and valuing robustness, both in the physical hardware and in the software engineering. The home introduces unknowns to any system, whether it be power outages, flaky networking, occupants that move sensors, and so on. The simple fact that the system will be running constantly, instead of being restarted each day (as it might have been during the development phase), can introduce software and hardware problems, such as memory leaks or "disk full" errors. The equipment not only has to run the actual deployment system, but it also needs spare capacity to cope with the overhead of being part of a research system. These overheads include logging for subsequent analysis, both in the device itself and through a cloud service (preferably) for offsite backup; remote access and control so the system can be monitored and fixed without site visits when possible; and debugging and development tools so the system can be modified in place without relying on external software or hardware. It should also be flexible enough that if, after a pilot deployment, the focus of the study shifts toward some interesting and novel use, the equipment can cope with that change in focus and be repurposed to support a new study.

Another cost tradeoff is deciding whether to use off-the-shelf hardware or build custom hardware. There are several reasons for using off-the-shelf systems when possible. Compared to designing, implementing, and manufacturing custom hardware, off-the-shelf hardware is easier and faster to acquire. Additionally, off-the-shelf hardware is often cheaper than custom hardware due to economies of scale, even just considering the bill-of-materials cost, let alone the development time. This option also allows flexibility in the number of devices to be deployed. Often, adding a new off-the-shelf device requires only a little time and small delivery fees, while trying to build one custom device can require reordering multiple base components, setting up the assembly space, and finding time to build and test the new device. Finally, using off-the-shelf hardware frees the project team from either hiring or learning skills such as soldering, PCB design, and industrial design.

However, off-the-shelf systems are limiting; they may do certain things very well but permit only limited reprogrammability or configurability, and, often, research systems may wish to go beyond that envelope by the very nature of the work. Off-the-shelf systems may not have sufficient robustness or logging capabilities required for research analysis. Frequently, device functionality is limited by a desire from the manufacturer to keep the costs down. Even with these drawbacks, given the time required to develop and debug custom hardware, we recommend using off-the-shelf hardware whenever possible and developing custom hardware only when necessary or if the device itself is part of the research focus.

Finally, a cost tradeoff exists between time before the study and time during the study. Consider a project that uses experience-sampling surveys to assess participants' opinions about a deployed prototype. This could be accomplished by either automating the process using a custom software program written and deployed by the research team, or manually emailing a questionnaire to participants each week/day/hour as appropriate. Depending on the size of the study and the complexity of building out a new tool, either may be appropriate. Carefully and realistically evaluating the time requirements of each option will help select the best approach. Sending a preformed email message to ten people each day for two weeks is certainly easier than authoring a custom application that has to run reliably on their hardware for two weeks. On the other hand, if the number of participants were a hundred and the timescale were months, it would perhaps be well worthwhile to write the custom application.

Sometimes, time can be traded for money, for example, by using Amazon's Mechanical Turk to process data, or by choosing a more expensive hardware/software system that does what is wanted "out of the box" rather than using a cheaper system that requires significant tweaking and modifying before deployment.

> **LESSON**
>
> In-home studies offer many ways to ask the same basic research question, with cost tradeoffs in terms of money, time before the study, and time during the study. Carefully consider which is best, and value robustness and flexibility for the equipment deployed, as the home environment will cause unforeseen issues.

VISITING THE HOME

Our final suggested consideration for all in-home studies concerns the frequency and timing of site visits. Scheduling times to visit homes can be difficult, especially if the study design requires talking with multiple family members. Be prepared to be available in the evenings and on weekends. We have found that offering to bring a pizza or other dinner items can be a valuable icebreaker and allows for a meeting during the dinner hour. Always send a reminder email, or make a reminder call, a day or so ahead, or risk showing up and finding the home empty (not a fun experience). Ideally, minimize the number of times required to visit the home, for example, by visiting only at the beginning and end of the study. If additional data collection is required during the middle of the study, try to do it through email or phone calls.

For safety reasons, it is important to avoid visiting an unknown participant's home by oneself. Particularly on the first visit, always bring at least one other person. While we have never had a safety incident, there have been a few visits where it felt more comfortable to have multiple people visiting together. In addition, since one person will be focused on interacting with the members of the household, having a second person will help ensure that required steps are not forgotten. Also, they can take additional notes as needed.

In very rare cases, it may be possible to conduct an in-home deployment without visiting the home. This is worth considering because it becomes possible to work with families that are outside of the researcher's immediate geographic region. For the SPARCS study (Brush, Inkpen, and Tee 2008), we deployed a photo- and calendar-sharing prototype to pairs of families (e.g., grandparents and kids) to run on their home computers. We sent the software by email and then did phone support with families to get it installed. The prevalence of video-calling programs such as Skype makes it easier to consider studies where technology is sent to participants, and then they are interviewed remotely. (For more on this approach, refer to Chapter 2 on conducting remote interviews with participants.) However, in most cases, the researcher will need to go to the house to deploy the system and possibly also visit after the study to gather data on how the system was used (e.g., photographs or interviews) and/or to tear down the deployed system.

LESSON

Treat home visits with care. Be well prepared in order to maximize the utility of each visit (e.g., pilot the visit and send reminders), expect scheduling to be difficult, and do not go alone if the participants are strangers.

PREHEAT STUDY AND METHOD

We now focus on our case study, the in-home deployment of our PreHeat home heating system. We cover the goals of our deployment, why we conducted an in-home deployment, and what data we collected and analyzed. Using occupancy sensing and a novel prediction algorithm, PreHeat strives to reduce the energy consumption of a household's heating system without compromising the thermal comfort of household members. PreHeat heats a space when it is occupied and also builds a predictive model so that even when a space is not currently occupied it may be proactively heated in advance of the next predicted occupancy. This allows PreHeat to react dynamically to the occupancy patterns in a household on a given day.

The two key metrics in evaluating whether PreHeat performs better than commonly used methods for heating homes are the amount of energy used to heat the home and the MissTime (time in which an occupied space is not within 1°C of the predefined temperature setpoint). MissTime is used as an indicator of thermal comfort. We wanted to compare the PreHeat algorithm against two other conditions: using a heating schedule (Schedule) or leaving the thermostat set at a single temperature (AlwaysOn).

One approach to quantifying energy used and MissTime for the three conditions (PreHeat, Schedule, and AlwaysOn) would have been to use a simulator (such as EnergyPlus), an approach some previous research has adopted (Gupta, Intille, and Larson 2009; Lu et al. 2010; Mozer, Vidmar, and Dodier 1997). However, we had concerns about how accurate a simulation would be and whether the simulation would capture the intricacies of a real-world deployment, and in any case, we needed to deploy sensors in real homes to get the fine-grained occupancy data (at the room scale) for the PreHeat algorithm. This led us to consider a full prototype deployment.

Several factors enabled us to make the decision to go ahead with this deployment. We had conducted several lower-overhead studies deploying GPS receivers (Krumm and Brush 2011) and temperature sensors (Scott et al. 2010), so we were confident that our deployment would have good results. We could limit the overhead of deployment by using our own homes and those of close collaborators, which also simplified, though by no means eliminated, privacy concerns. Based on our work on the Microsoft .NET Gadgeteer rapid prototyping platform, we had the expertise to use custom hardware, which gave us full control of the system behavior right down to the hardware level. Feeling comfortable that we could fix issues that arose "in the wild," we decided to go ahead with performing actual control of home heating systems to evaluate PreHeat. We believed this offered a clearly superior evaluation technique to other options.

We conducted our deployment of PreHeat in five family homes: three were our own homes and two were homes of colleagues working in our research lab. These homes were also spread across two countries, with three homes in Seattle, Washington, USA, and two in Cambridge, UK. We chose these homes based on a number of considerations. First, because the study involved replacing a critical piece of infrastructure—the home's heating system—we wanted to ensure that there was at least one computer expert in each house who could work with the system in case of emergency. Second, because the homes were our own and our colleagues', we could visit the homes more freely to debug and fix the systems as necessary, and the participants were more open to living with an experimental system because they trusted us to fix any problems quickly. Finally, we selected homes where participants left the house for significant parts of the day, so there was the potential for energy savings by identifying those time periods. Because our evaluation took place in our own homes and those of close colleagues, we explicitly restricted ourselves to collecting quantitative data and excluded any qualitative data to avoid the potential that we might be biased toward our own system. Chapter 8, "Autobiographical Design in the Home," describes advantages, disadvantages, and things to consider when using one's own home as a deployment site.

We evaluated two other homes of colleagues who expressed an interest in being part of the study as potential deployment sites, but excluded them. We excluded one UK home because we could not easily augment the gas meter to be electronically readable (so we would not have been able to gather energy use data easily), and one US home because the home was occupied nearly all of the time and therefore offered few opportunities to reduce heating time.

Our study had three phases. In our Debug phase, which lasted around seven days in each house, we replaced the home's heating system with our custom system, initially using the Schedule condition, which was most similar to the occupants' previous thermostats. In the three US homes, we controlled whole-house forced air heating systems by deploying a custom thermostat (Figure 9.1) and a home heating server that ran the heating algorithms, sent commands to the thermostat, and logged data. For sensing when occupants were home, we used active RFID (radio frequency

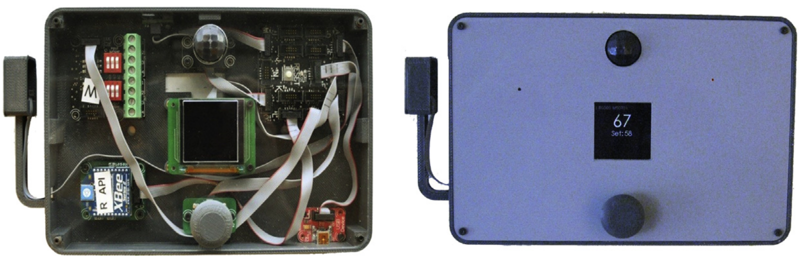


FIGURE 9.1

.NET Gadgeteer hardware for US thermostat.

FIGURE 9.2

To sense when participants were home we placed small, active RFID tags on their key chains and attached an RFID reader to our home heating server.

identification) tags that occupants put on their key chains (Figure 9.2) and a reader attached to the home heating server. In the two UK homes, we additionally augmented hot water radiators and underfloor heating to exercise per-room control. In these homes, we sensed per-room occupancy using passive infrared motion sensors (Figure 9.3). The installation was particularly complex in the UK houses

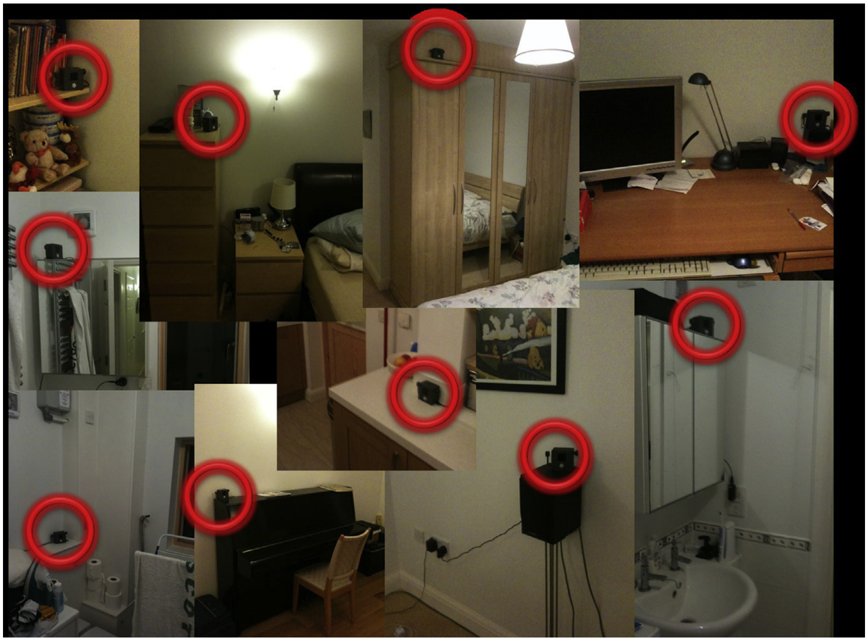


FIGURE 9.3

Room units deployed all over a house in the UK.

because each room had to be separately instrumented. This required installing and configuring around fifteen different hardware elements per house, and, for some elements, rewiring 240V electrical circuits. Naturally, a system of this complexity did not work flawlessly the first time, and we needed several return visits to refine the prototype. Starting with our own houses first reduced the overhead of making system changes. The Debug phase also gave households a chance to adjust the temperature setpoints if desired, and to fine-tune the heating schedule to allow the Schedule condition to match their actual occupancy pattern as much as possible. This also mitigated any potential novelty effect common with new technology.

Phase 1, the first data collection phase, lasted fourteen days, and our system alternated daily between using AlwaysOn and Scheduled algorithms to heat the homes. During this phase, the system collected fourteen days of occupancy data that was then used by the PreHeat's prediction algorithm in the next phase. The final and main phase of the study, Phase 2, alternated between Scheduled heating of the home and PreHeat's predictive heating algorithm. This phase lasted at least forty-eight days in every house. Using data from Phase 1, we estimated what the energy use of AlwaysOn would have been during Phase 2 to make a final three-way comparison.

In comparing the performance of the three heating algorithms, PreHeat, AlwaysOn, and Scheduled, there were two main confounds to overcome: weather and house occupancy variability. To cope with variable weather (most importantly, the variation in outside temperature both day to day and as winter became spring), we alternated which algorithm was running on a day-by-day basis. To try to account for and balance out potential schedule differences that might unfold in each house, we ran the study for many weeks. Taking Phases 1 and 2 in total, deployments ranged from sixty-two to eighty-four days.

Our results showed that PreHeat enabled occupants to make more efficient use of home heating (i.e., achieve a better tradeoff of energy used versus MissTime incurred), particularly in UK houses with per-room control, where energy savings of 8 to 18 percent were achieved over Scheduled, and 27 to 35 percent over AlwaysOn, while still reducing MissTime by 38 to 60 percent.

CHALLENGES AND LESSONS LEARNED

We now describe the main challenges we had to overcome when conducting the PreHeat study and the lessons we learned from dealing with them.

IMPLEMENTING THE MINIMUM VIABLE PROTOTYPE

For a system that controlled a vital part of a real house's infrastructure, we had to build out the system's features and user interface (UI) so that it largely matched the capabilities of occupants' existing heating systems. In the UK, for example, this meant including hot water control and scheduling as well as heating control, and supporting per-room temperature setpoints. In the United States, we had to deal with controlling the circulation fan as well as furnace control. However, while these are non-negotiable features, we were also aware of many other features (e.g., remote control of heating) that would have been nice to have, but were not strictly necessary for our particular study. To strike a balance on what to implement, we found a useful analogy in the advice popularized by The Lean Startup Movement (Ries 2011), that start-up companies must focus on implementing the Minimum Viable Product (MVP) and eliminate all extra features so they can ship their first product rapidly. In this vein, we tried to carefully identify and implement the minimum viable prototype of a heating and occupancy sensing system. This helped make possible the development effort to build the prototype in a reasonable period with the resources we had.

We identified some features that were required as part of a minimum viable prototype, for example, the ability to manually override the setpoint temperature with a simple action on each device. However, we decided that other existing features of heating systems could be acceptably implemented by participating households sending an email to us, the researchers, rather than as a functioning UI in the prototype. For example, changing the temperature setpoints and activating vacation mode were accomplished by us manually editing the settings files when participating households emailed us asking for these changes in the system. Remotely editing the configuration files was much simpler than implementing an end-user-facing control UI.

Piloting in our own homes and during the debug phase was critical for identifying whether we had missed any required features and correcting problems. For example, we had not initially included mechanisms to dim the screens of the UK devices at night, but as soon as we deployed them, we received feedback that glowing screens were disturbing at sleep time, and we added an automatic dimming feature.

LESSON

Think carefully about what set of features make up the minimum viable prototype and where development time and effort can be saved. Only automate features that would otherwise involve excessive researcher time or occupant inconvenience to update manually.

DEPLOYING A SAFETY CRITICAL SYSTEM

In our deployment, we controlled the heating of the home, a safety-critical system, particularly in the winter months. We were very concerned about making sure that the system worked robustly and, if any problems occurred, there were fallbacks in place at every level, from the prototype hardware device up through the software layer. Our guiding principle was to always "fail safe" and keep the house at a comfortable temperature if a problem occurred.

For example, if a device controlling the heating ever lost communication with the computer responsible for issuing the heating commands, the device defaulted to the occupants' preferred setpoint so that the house would stay warm even though a failure occurred. Another example is that all of the various hardware and software elements in the system were built with "watchdog" timers, that is, systems that monitor whether an element is operating correctly and will restart or reset it if not. This was true of our home heating server software (where a separate program constantly monitored whether the heating program was running and restarted it if not), home heating server hardware (the PC BIOS was set to automatically turn back on if there was a power cut), the firmware in our embedded devices (which rebooted and reset the hardware if the system was not functioning correctly), and of the communications between them (our network protocols included heartbeat signals and acknowledgments of every command so that the home server could monitor the status of the embedded devices at frequent intervals).

Such care is rarely necessary for in-lab studies or demonstration prototypes, which have to operate for a few minutes or hours only. For real-life deployments, this care offers a valuable safety net for rare or hard-to-anticipate errors; for example, electrical noise from the furnace occasionally caused one of our embedded device processors to hang and require rebooting.

To cope with cases where these automatic safeguards could not recover the system, we implemented an alerting mechanism, specifically, "emergency" state notification emails that were sent every thirty minutes whenever the system was not fully operational. In addition, we logged the times the system was in an "error" state so we could account for that during our data analysis phase.

By using watchdogs, heartbeats, acknowledgments, and alerts, we achieved a 99.8 percent uptime during the study, and our system "failed safe" at other times. For nearly all problems, the system recovered automatically and never reached the threshold where an email alert was sent that required manual attention. Most failures ended up being due to contention in the 2.4GHz radio band, in which our system was competing with WiFi devices; however, these were temporary, and the system recovered automatically. In case all this work sounds like overkill, remember that we were replacing a critical system for which one normally expects to have a 100 percent uptime.

LESSON

An in-home deployment requires the system to run extremely robustly, or there is the risk of having participants respond more to the quality problems rather than the system being evaluated. Think about techniques for automatic recovery from unpredicted errors, "fail-safe" behavior, and automatic alerts should unrecoverable failures occur.

SENSING IN THE REAL WORLD IS TRICKY

In the PreHeat study, we needed to sense the following real-world variables: the temperature of the space, the occupancy of the space, whether or not the heating system was turned on in each space, and how much natural gas was being used. Since we were in control of switching the heating system on and off, we could record that information directly. That left the natural gas reading, the temperature, and the occupancy variables that needed to be sensed in real time.

To sense gas usage, we used different methods in the United States and UK because of having to adapt to differences in local infrastructure. In the United States, the local utility company made gas meter data available from its smart meters. In the UK, we augmented the gas meters with off-the-shelf devices from RFXCOM that detected the meter's numeric digits rotation, and hence gave us gas readings.

We discovered that sensing the temperature was not as easy as it might appear. When validating the temperature sensor's readings from our prototype device against an accurate lab thermometer in our early device designs, we found that heat generated by the processor in the device was measurably affecting the temperature readings. Therefore, we ended up trying several case designs, finally settling on one that had the temperature sensor on a separate curved arm to maximize thermal isolation (see Figure 9.1).

To sense space occupancy, we bought several off-the-shelf systems and did test deployments to evaluate them relative to our needs. Our first attempt was to use the household WiFi router as a way to detect when smart phones entered or exited the house. Testing this approach in one of our homes revealed that this approach did not have sufficient granularity (often the phones would not connect to the WiFi for up thirty minutes after arriving), nor was it reliable (the phones frequently went to sleep and appeared to have left the house). Next, we tested putting off-the-shelf motion sensors in each room. While these devices were inexpensive, they were often unreliable when run continuously for several days; for example, a radio packet loss would cause us to lose a "departure" event, and so a large time period would be misclassified. We finally located an active RFID tagging solution that was a reliable way of doing house-wide occupancy detection. Tagging did come with a cost to our participants, as they were required to carry the tag itself on their key rings; however, we felt that requirement could be managed. Figure 9.2 shows the tags we used. New options for tagging have been developed since our study, and if we were re-running the study today, we would likely use Bluetooth Low Energy (BLE) tags, which can be lower power and smaller than the Active RFID system we used.

While Active RFID tagging solved the issue of household-level occupancy, in the UK, deployments were still needed for per-room occupancy. This was accomplished by having our custom prototype control unit for each space contain a motion sensor. By using careful placement of the units and tuning the parameters, we found that we could get the accuracy we needed in our test houses.

In each case, we needed to run multiple trials against known ground truth in order to get a reliable and accurate system. In the case of the occupancy, we had to deploy the systems in our houses and manually record our movements to compare against the sensor data and verify that the system was working as expected.

LESSON

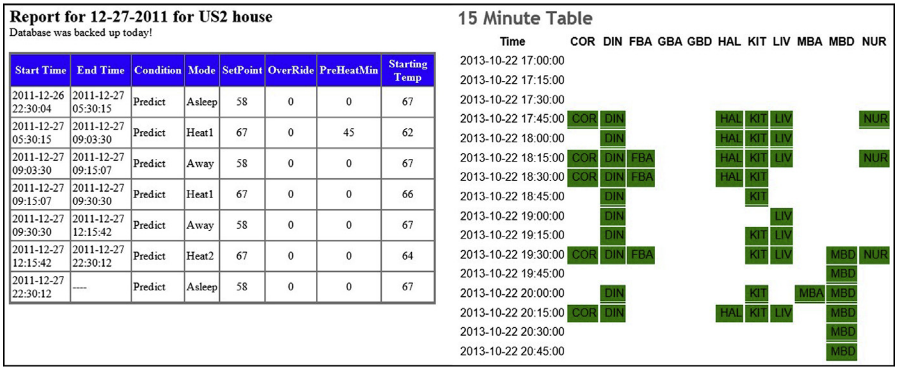
Sensing in the real world is difficult, and more than one sensing solution may be needed if there are differences in the households. Be sure to validate the sensed data to ensure the sensors are accurate, and plan for multiple iterations of prototyping in case the first solution is not robust.

ENSURING VALIDITY OF THE COLLECTED DATA

A study is all about collecting data, and thus the most important thing a researcher can do is ensure that the data collected is both accurate and preserved in case of system failure or downtime. We stored local log files so that any network connectivity loss or unexpected power events would not result in lost data. All data was eventually uploaded to an SQL database, and that database was backed up each day, both to the cloud and to a local solid-state USB drive. Thankfully, we did not experience any serious hardware issues, and our backups were not required, but several times, power failures or other events occurred that had the potential to have caused data corruption.

The care we took in system development allowed us to respond quickly to failures (e.g., devices unplugged) and avoid losing days from the study. Moving beyond system robustness, we also had to ensure that we were collecting valid data throughout the study. In particular, we were concerned about ensuring the validity of the home occupancy data. Both occupancy-sensing methods we used had the potential for problems. The active RFID tags on people's key chains, which tracked whether they were home or not, had a relatively limited range, requiring participants to put their keys within the range of the reader attached to the home heating server. The passive IR sensors used for room-level occupancy in the UK had trouble at night, when people moved much less while they were sleeping, and also had trouble covering some of the larger rooms.

To carefully monitor the occupancy data being collected and quickly spot any potential problems, we wrote a program that each morning emailed a visualization of the previous day's occupancy data for each household. Figure 9.4 shows an example of the information sent daily, including the success of data backups, the heating schedule that was followed in the house, and occupancy data, including motion sensor (US houses had a single motion sensor in the thermostat) and RFID data. At least one project team member inspected each visualization every morning and followed up as necessary with household members for clarification while the events were still fresh in their minds. In addition, we ran consistency checks examining the recorded data in the United States for situations when the RFID tag disappeared and reappeared frequently—which were typically RFID radio range problems—and in the UK, comparing the RFID data to the room-level motion data to highlight time when the house was occupied but the motion sensors were not



Report for 12-27-2011 for US2 house
Database was backed up today!

Start Time	End Time	Condition	Mode	SetPoint	OverRide	PreHeatMin	Starting Temp
2011-12-26 22:30:04	2011-12-27 05:30:15	Predict	Asleep	58	0	0	67
2011-12-27 05:30:15	2011-12-27 09:03:30	Predict	Heat1	67	0	45	62
2011-12-27 09:03:30	2011-12-27 09:15:07	Predict	Away	58	0	0	67
2011-12-27 09:15:07	2011-12-27 09:30:30	Predict	Heat1	67	0	0	66
2011-12-27 09:30:30	2011-12-27 12:15:42	Predict	Away	58	0	0	67
2011-12-27 12:15:42	2011-12-27 22:30:12	Predict	Heat2	67	0	0	64
2011-12-27 22:30:12	----	Predict	Asleep	58	0	0	67

15 Minute Table

Time	COR	DIN	FBA	GBA	GBD	HAL	KIT	LIV	MBA	MBD	NUR
2013-10-22 17:00:00											
2013-10-22 17:15:00											
2013-10-22 17:30:00											
2013-10-22 17:45:00	COR	DIN				HAL	KIT	LIV			NUR
2013-10-22 18:00:00		DIN				HAL	KIT	LIV			
2013-10-22 18:15:00	COR	DIN	FBA			HAL	KIT	LIV			NUR
2013-10-22 18:30:00	COR	DIN	FBA			HAL	KIT				
2013-10-22 18:45:00		DIN					KIT				
2013-10-22 19:00:00		DIN						LIV			
2013-10-22 19:15:00		DIN					KIT	LIV			
2013-10-22 19:30:00	COR	DIN	FBA				KIT	LIV		MBD	NUR
2013-10-22 19:45:00										MBD	
2013-10-22 20:00:00		DIN					KIT		MBA	MBD	
2013-10-22 20:15:00	COR	DIN				HAL	KIT	LIV		MBD	
2013-10-22 20:30:00										MBD	
2013-10-22 20:45:00										MBD	

FIGURE 9.4

We set up the system to send emails each morning (illustrative excerpts above) showing what happened in the house the previous day. These assured us that data was being backed up and allowed us to look for any issues in the heating control periods and occupancy data.

active. During the study, we encountered both "social" issues, for example, where someone forgot their keys, and "technical" ones, such as RFID radio range problems. Using these cross-checks, we manually identified fifty-seven errors in the sensed occupancy data that could affect the analysis. To ensure our results (e.g., the MissTime calculation) were accurate, we created a "ground truth" occupancy table for use during the data analysis.

> **LESSON**
>
> Remote monitoring allows the researcher to keep a close eye on the data being collected so he or she can have confidence in its validity and address any issues quickly, for example, by asking participants while their memory is fresh. Log everything, in case it is useful, in backed-up storage, and if possible, log complementary data sources that can be used to cross-validate each other. It may be necessary to manually correct some data errors to maximize the validity of the analysis.

CUSTOM HARDWARE REQUIRED

In order to build out the PreHeat system, we first looked for off-the-shelf hardware and software systems to accomplish what we wanted, but we ended up building some custom hardware and software in cases where the available off-the-shelf systems was inadequate for one reason or another. In addition to developing the custom software for the heating control algorithms, which was the main focus of the research, we identified three areas where we needed to source and purchase hardware for the house and started looking for off-the-shelf solutions. First, we needed a device to run the heating control software. Second, we needed full control over the heating system. Third, we needed to sense the temperatures, occupancy, and gas readings as described in the previous subsection.

For the home heating server, we elected to use an off-the-shelf small form factor PC. While this was more a powerful and expensive device than required to run a heating system, it allowed us to run the development tools on the devices themselves, facilitating in-situ debugging, and making remote control simpler (e.g., via remote desktop).

For controlling the heating system, our investigation discovered that there were no off-the-shelf devices that provided the fine-grained control (for each of our diverse heating systems) and logging that our system would need in order to test our algorithm. While we did find off-the-shelf radio-controlled thermostatic radiator valves (TRVs) that could perform per-room heating control in the UK, we realized we needed to build custom radio transmitter units to control them. Thus, we ended up designing three types of custom devices. In the UK, we had a device in each room that performed occupancy and temperature sensing and local TRV control, and provided our custom UI. We also had an additional type of device that simply actuated the boiler through 240V relays. In the United States, we built a device replacing the wall thermostat, which performed temperature sensing, supplied a custom UI, and actuated the furnace.

We did not start from scratch to build these custom devices, but instead based our prototypes on Microsoft .NET Gadgeteer (Villar et al. 2012), a modular prototyping toolkit. While we had to design a few custom hardware elements (e.g., the transmitter that communicated with the thermostatic radiator valve actuators), most of the hardware was standard .NET Gadgeteer modules. This saved us a lot of time over building fully custom hardware from scratch. Being a modular toolkit, there was also a lot of overlap between the hardware in the two types of UK unit and the US unit—indeed, they actually run the same software that detected the unit type on the fly by detecting the presence of various hardware modules. The inside of a US thermostat unit is shown in Figure 9.1, illustrating the modular hardware. We used a 3D printer and laser cutter to manufacture custom device cases.

LESSON

Use off-the-shelf hardware and software where possible. If custom hardware is needed, rapid prototyping toolkits can reduce the total workload and enable faster design iterations.

CONSISTENCY ACROSS STUDY CONDITIONS

We conducted a study with three heating conditions (AlwaysOn, Scheduled, PreHeat) that were each used on different days depending on which phase of the study the house was in. To ensure household members behaved in the most consistent manner, for example, doing a heating override if they were cold, we did not want people's interactions with the system to change when the heating condition changed. Such consistent behavior is important to ensure the validity of the results. We therefore took several steps to ensure that the user experience across the study conditions was as uniform as possible.

First, we developed a UI and system configuration that hid the existence of different conditions. Specifically, while the device display informed a user of the current setpoint of the system, it did not specify the mode of operation, nor why the current heating setpoint was chosen. Second, if the participant decided to override the current temperature they had the choice of doing so "until the next sleep" (the time that the system went to the nighttime set point) or "until the next wake" (the morning time when the system went to the daytime set point). Since these two times were the same in every condition, the participant had no feedback on which condition was being evaluated at that time. Finally, our day-by-day alteration of the study condition, primarily done to mitigate the effect of weather, also hid the current condition from participants better than having each condition be a large contiguous block of time.

LESSON

Think about how to implement study conditions so they present participants with a coherent experience across conditions as much as possible, to avoid measuring participants' reactions to irrelevant differences.

THE END OF THE STUDY

At the end of our study, we allowed households to keep the system if they wished because they could maintain it themselves. Three houses opted to continue running PreHeat. We also shared the output of the study with the participants. We advise doing the same whenever possible, as many people participate in deployments because they are interested in research and the findings may be more valuable to them than any other gratuity.

One US house appreciated the reactive heating on the weekends, which kept the house warm when they were home. Eventually the husband wanted the system removed when he was going to host his office party because of a power cord that was unattractively taped across the ceiling to power the thermostat. (Attractiveness of the deployed system does matter, but was not part of our minimum viable prototype.)

In the two UK houses, the system continues to run today, more than four years after the original study. It requires occasional maintenance (e.g., the disk on the home PCs fill up because of log backups), and some features that were excluded from the minimum viable prototype were subsequently implemented for long-term use, such as a "master off" for summer. Additionally, we showed the lead occupant how to change the setpoint in the configuration files to replace the "email us" interface we had used during the study.

LESSON

Make a plan about how the study will end, and either allow people to continue using the prototype or smoothly return their homes to the pre-study state.

CONCLUDING REMARKS

In-home deployments result in rich data that would be impossible to get from any other evaluation method. While all researchers should carefully consider when an in-home study is most appropriate, once the decision is made to move forward with an in-home deployment, there are many steps one can take to ensure that deployment is successful. The best chance of collecting good data is had by carefully considering the functionality necessary for a minimum viable prototype, evaluating a range of technical choices—including both off-the-shelf and custom hardware—and allowing plenty of time to debug and pilot the study. Remote monitoring will alert the researcher to any potential problems and provide time to correct them, while making plans for how to handle privacy concerns and how to end the study will help avoid any surprises.

Our experience deploying PreHeat and other in-home studies and, in particular, the amount of custom engineering work these studies required, was part of the motivation to develop and open source both the .NET Gadgeteer (Villar et al. 2012) and the Lab of Things platforms (Brush et al. 2013). We hope these platforms will allow future studies to reuse and build upon past engineering efforts.

.NET Gadgeteer (http://netmf.com/gadgeteer/) is an open source modular prototyping platform that supports solderless assembly of devices through modular hardware and standard cables, enables software development and debugging using managed C# code and Visual Studio, and facilitates device case design through standard mounting fixtures on each module. Between the various manufacturers making .NET Gadgeteer-compatible hardware, there are currently more than one hundred types of .NET Gadgeteer modules available for purchase.

The Lab of Things SDK (http://lab-of-things.com/) strives to reduce the amount of development necessary to conduct a home study and, more generally, grow a community of people conducting studies that contribute to shared infrastructure. We hope that by working together we can ease the burden for everyone and enable in-home deployments at much larger scales than typically occur—hundreds of homes rather than tens—and with greater geographic diversity.

We hope these two toolkits are beneficial to researchers working in the home, and that the more general lessons we learned and shared in this chapter provide useful guidance for home deployment studies.

REFERENCES

Brush, A.J., Inkpen, K., Tee, K., 2008. SPARCS: Exploring Sharing Suggestions to Enhance Family Connectedness. Conference on Computer supported cooperative work (CSCW 2008). ACM Press, New York, NY.

Brush, A.J., Filippov, E., Huang, D., Jung, J., Mahajan, R., Martinez, F., Mazhar, K., Phanishayee, A., Samuel, A., Scott, J., Singh, R.P., 2013. Lab of Things: A Platform for Conducting Studies with Connected Devices in Multiple Homes. UbiComp (Adjunct Publication), 35–38.

Brush, A.J., 2010. Ubiquitous Computing Field Studies. In: Krumm, J.C. (Ed.), Ubiquitous Computing Fundamentals. Chapman and Hall/CRC.

Brush, A.J., Inkpen, K., 2007. Yours, Mine and Ours? Sharing and Use of Technology in Domestic Environments. In: Proceedings of the 9th International Conference, Ubiquitous Computing (UbiComp 2007).

Brush, A.J., Inkpen, K., Johns, P., Meyers, B., 2011. Speech@Home: An Exploratory Study. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI). ACM Press, New York, NY.

Dix, A., Finlay, J., Abowd, G., Beale, R., 2004. Human-Computer Interaction, 3rd ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Gupta, M., Intille, S.S., Larson, K., 2009. Adding GPS-Control to Traditional Thermostats: An Exploration of Potential Energy Savings and Design Challenges. In: Proceedings of the Pervasive. Cambridge, MA.

Hnat, T., Srinivasan, V., Lu, J., Sookoor, T., Dawson, R., Stankovic, J., Whitehouse, K., 2011. The Hitchhiker's Guide to Successful Residential Sensing Deployments. In: Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems. Nov. 2–4, 2011. ACM Press, New York, NY.

Kay, M., Choe, E.K., Shepherd, J., Greenstein, B., Consolvo, S., Watson, N.F., Kientz, J.A., 2012. Lullaby: A Capture & Access System for Understanding the Sleep Environment. In: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (Ubicomp '12). ACM Press, New York, NY.

Krumm, J., Brush, A.J., June 2011. Learning Time-Based Presence Probabilities. In: Pervasive. Springer Verlag, Berlin, Heidelberg.

Lu, J., Sookoor, T., Srinivasan, V., Gao, G., Holben, B., Stankovic, J., Field, E., Whitehouse, K., 2010. The Smart Thermostat: Using Occupancy Sensors to Save Energy in Homes. Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys, '10). ACM Press, New York, NY.

Mozer, M.C., Vidmar, L., Dodier, R.M., 1997. The Neurothermostat: PreHeat Optimal Control of Residential Heating Systems. In: Advances in Neural Information Processing Systems 9. MIT Press, Cambridge, MA, pp. 953–959.

Pousman, Z., Romero, M., Smith, A., Mateas, M., 2008. Living with Tableau Machine: A Longitudinal Investigation of a Curious Domestic Intelligence. In: Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp 08). ACM Press, New York, NY.

Ries, E., 2011. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses [online]. Crown Business, Lean Startup. http://theleanstartup.com/.

Rogers, Y., Sharp, H., Preece, J., 2011. Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons.

Scott, J., Brush, A.J., Krumm, J., Meyers, B., Hazas, M., Hodges, S., Villar, N., Sept, 2011. PreHeat: Controlling Home Heating Using Occupancy Prediction. In: Proceedings of the UbiComp. ACM Press, New York, NY.

Scott, J., Krumm, J., Meyers, B., Brush, A.J., Kapoor, A., May 2010. Home Heating Using GPS-Based Arrival Prediction. In: Proceedings of Pervasive 2010 Workshop on Energy Awareness and Conservation through Pervasive Applications. Microsoft Research.

Villar, N., Scott, J., Hodges, S., Hammil, K., Miller, C., June 2012. .NET Gadgeteer: A Platform for Custom Devices. In: Proceedings of Pervasive 2012, Lecture Notes in Computer Science.

CHAPTER

Field Trials with Multiple Connected Households

10

Tejinder K. Judge*, Carman Neustaedter§

Google Inc., Mountain View, CA, USA; School of Interactive Arts and Technology, Simon Fraser University, BC, Canada §*

INTRODUCTION

Field trials are a method for studying the use of technologies by real end users in their naturalistic setting (Brown, Reeves, and Sherwood 2011). The goal of a field trial is to evaluate a system in either a formative or summative manner in which investigations typically go beyond exploring usability issues, such as task performance time, to understanding and exploring the deeper issues that end users face when using a technology amidst their everyday lives. Thus, field trials explore real usage practices along with the ways in which users assimilate technology into their everyday routines and context. Field trials can also generate information about the value of the technology in the lives of users. Given the real-world context of field trials, they are often thought of as a useful method for gathering longitudinal data about the usage of a system, in particular after any novelty effects have worn off (Nielsen 2004). Over the years, field trials have been referred to with a variety of names, including field deployments, field evaluations, field studies, in-situ studies, and "in-the-wild" studies.

In this chapter, we discuss the ways in which we have utilized field trials as part of our research on the design of video-based family communication systems. We have deployed a number of different systems for family use in the home and, in each case, learned valuable lessons about how to conduct field trials. This chapter focuses in particular on the case where collaborative technologies are deployed in domestic settings where they provide connections *between multiple* distance-separated homes. Thus, it is not a matter of testing out or trying a technology in a single home; instead, the usage of the technology spans several homes, where each is tightly coupled and interwoven with the practices that emerge in other homes. Field trials are challenging in and of themselves; however, we have found that when field trials move beyond just a single home to connect multiple homes with new prototype technologies, the challenges increase dramatically.

This chapter unfolds as follows. First, we provide background knowledge about field trials in which we describe the basics of the method along with pointers to examples of field trials within the Human-Computer Interaction (HCI) literature. Second, we focus on our own experiences of conducting field trials by describing

Studying and Designing Technology for Domestic Life. http://dx.doi.org/10.1016/B978-0-12-800555-2.00010-1

a system called Family Portals that we designed, developed, and deployed in families' homes (Judge et al. 2011). Family Portals is an always-on multi-family video media space that provides a continuous video connection between three different homes. Third, we outline the various lessons that we learned about conducting field trials through our own experience of deploying Family Portals. Lastly, we step back and reflect on the ways in which we think multi-location field trials can and should be conducted in domestic settings.

FIELD TRIALS IN HCI

Field trials involve the evaluation of a system that either is already in place within a setting or is given to users who are asked to use the system as a part of their everyday activities in their own environment (e.g., work or home). Field trials allow researchers to understand how similar technologies might be used if they were released commercially (Brown et al. 2011). Field trials can also lead to an understanding of user behavior and usage patterns that can be used to iterate on the design of a system or to develop design guidelines and recommendations for future similar systems.

In field trials, users are typically asked to use a system for a given time period as part of their daily routine while researchers collect data on this usage. A field trial can range from a few days to a few months or even years, depending on the nature of the research, the commitment of the users to using the system, and the frailty of the prototype being evaluated. Data from field trials are collected using a variety of methods including system logs, user interviews, observations, and self-report diaries. Prototypes tested in field trials can be fully functional versions of a final product or they can be early explorations that may have little to do with a final product (Korn and Bodker 2012).

Over the past two decades, field trials have become an increasingly common method for conducting research in HCI. In the first half of the 1990s, HCI research interests shifted from focusing on task performance in laboratory experiments to evaluating user interactions with technology and their understanding of the system (Olson and Teasley 1996; Barkhuus and Rode 2007). This is because traditional evaluation methods and metrics such as task completion times "… fail to capture the complexities and richness of the real world in which the applications are placed" (Rogers et al. 2007; Greenberg and Buxton 2008). Early field trials were investigations of prototypes that were used longitudinally in research lab settings. Examples of these include Portholes at EuroPARC (Dourish and Bly 1992), the Active Badge system at AT&T Laboratories (Harper, Lamming, and Newman 1992), and the media space at PARC (Bly, Harrison, and Irwin 1993). Although these early field trials were conducted by researchers in their own workplaces, their reflections from long term usage were valuable in "… forming an understanding of what happens when these technologies became more widespread" (Brown et al. 2011).

We have also seen an expansion over the years from the design and study of computing technologies solely for the workplace to increased interest in computing technologies designed for the home and domestic environments. This includes commercial products such as video chat technologies (e.g., Skype, Google+ Hangouts), home entertainment systems (e.g., Apple TV, Chromecast), and technologies that promote health and wellness (e.g., Wii Fit). A similar shift has also emerged in HCI research where, starting in the late 1990s, there has been increasing interest in studying technology design for families. This naturally involves studying the way technology is used *in situ* because it is often very difficult to replicate domestic life and domestic environments in a laboratory setting. Such real-world testing and evaluation is not easy and, because of this, some researchers have created "living labs," such as MITs PlaceLab (Intille et al. 2006) and Georgia Tech's Aware Home (Kidd et al. 1999), to simulate home settings for design and evaluation research. By controlling the environment, researchers hope to reduce the uncertainty found "in the wild" while retaining many aspects of a field trial. Although field trials in a living lab can provide both quantitative and qualitative data, they tend to lack ecological validity (Creswell 2003) because study participants are asked to live in a new environment and may or may not be able to recreate their typical routines and lifestyles.

Given the limitations of laboratory-based evaluation and living labs, field trials that occur in real users' homes remain a popular method for evaluating domestic technologies and understanding their usage. In fact, we have seen many examples of the use of field trials in domestic computing research over the years. For the purposes of this chapter, these generally fall into two categories: field trials in single home units and field trials of technologies that connect people across multiple homes.

The first type, single-home or single-family field trials, involves individuals in a family using a system with other co-located family members. For example, Neustaedter and Brush (2006) evaluated a digital family calendar called LINC in the homes of four different families to understand their use of the technology over a period of four weeks (Neustaedter and Brush 2006; Neustaedter, Brush, and Greenberg 2007). Family members within single households used the calendar as a group, but interaction did not span multiple homes. Other single-home field trials include the evaluation of Family Archive, a multi-touch tabletop system that acted as a technology probe to learn about families' photo archiving practices (Kirk et al. 2010), and Froehlich et al.'s (2012) field trial of novel eco-feedback displays that enable families to learn about their water consumption. Chapter 9 provides more detail on planning and conducting single-home field trials

The second type of field trial is a multi-home or multi-family field trial. In this situation, people in multiple homes use the same system where it connects household members across locations. The simplest form involves dyads, where two households connect through the use of the system. An early example is the Casablanca project, in which multiple prototypes and early concepts, such as the Commute-Board and MessageBoard, were initially evaluated by researchers in their own homes, followed by field trials with actual users (Hindus et al. 2001). In these cases, the technologies were used to transmit information such as messages between households. Another example comes from Digital Family Portraits (Mynatt et al.

2001; Rowan and Mynatt 2005), a prototype technology in the form of a photo frame that shares awareness information between an older adult's home and her adult child's home. Digital Family Portraits was deployed between two homes for a period of one year to understand longitudinal usage. Findings from the field trial led to refinements in the design and an understanding of how the device was used to support awareness between families (Rowan and Mynatt 2005). Other examples include dyadic field trials of SPARCS, a system for sharing photos between two households (Brush, Inkpen, and Tee 2008), and our own Family Window, an always-on video media space used for connecting two distance-separated households (Judge, Neustaedter, and Kurtz 2010).

The above examples all focus on connecting two households; however, multi-home field trials can also involve more than two homes. For example, field trials of videoProbe connected a triad of households through the use a photo sharing system (Conversy et al. 2003), while Wayve connected small networks of families and friends (quartet, triad, dyads, and single households) using a messaging device (Lindley, Harper, and Sellen 2010). Although the difference between a field trial with two connected homes and a field trial with three or more connected homes seems like a trivial increase in locations, the difference is actually quite complex.

First, with three or more households, researchers must deal with more users and more dynamic relationships within and between the households, all of which could affect usage and adoption of the system. For example, in a field trial with two homes that have four members per household, there are eight users in total and twenty-eight unique relationships within and between households (i.e., six unique relationships per household and sixteen unique relationships between households). These relationships include husband-wife, father-child, child-sibling, and child-aunt, among others. In contrast, in a field trial with three homes that have four members per household, there are twelve users and sixty-six relationships within and between households. This is a much more complicated scenario for researchers to understand and study. Our own experiences, as well as that of others, have demonstrated that more homes and more family members connected to each other between homes creates additional technical challenges and complexities, as compared to studying a system in only one home location (Hindus et al. 2001; Plaisant et al. 2006; Judge et al. 2011). This is not to say that single-home field trials are not complicated in their own way, yet they are certainly different from a multi-home situation.

Next, we dive more deeply into exploring multi-home field trials by reflecting on our own field trial of the Family Portals system (Judge et al. 2011).

THE DESIGN OF FAMILY PORTALS

The Family Portals (FP) project was part of a larger series of explorations on the design of communication systems for families (Judge 2011). The overarching goal of these explorations was to understand how we could design video communication systems for families by changing the design paradigm from one involving "calling" family members (akin to commercial video chat systems available at the time) to one involving connecting homes with persistent video feeds. We did this by exploring the idea of "always-on" video in the home. Here the intention was that a family could leave a video feed running all the time between their home and that of remote family members. People in both households could glance every now and then at their "always-on video device" (e.g., a digital picture frame or tablet showing the video link) to see what people at the remote home were up to. They could also move from this awareness to interaction by communicating through features in the system or using other technologies, such as the phone, to augment the video feed.

We started by studying the ways in which people used existing commercial video chat systems (e.g., Skype, Apple FaceTime, Google+ Hangouts). This showed us that some families left video chat systems running for extended periods of time in order to share everyday life with a remote home (Judge and Neustaedter 2010). This was particularly prominent in families with children. Building on this idea, we created an always-on video system called the Family Window to connect two homes with a persistent video link (Judge et al. 2010). Chapter 8 describes our autobiographical design and evaluation of the Family Window. Following this, we conducted a field trial of the Family Window and found it was quite successful at connecting family members over distance: non-collocated family members valued the ability to see one another, and privacy concerns were minimal (Judge et al. 2010). Next, we decided to extend our design idea to explore how a triad of homes might connect with always-on video; here we wanted to try to "push the limit" to see when the idea of always-on video might break down or if it could continue to be extended to connect even more households in a positive way. This led us to the design of Family Portals, described in its entirety by Judge et al. (2011) and Judge (2011).

Next, we provide a brief description of the design of Family Portals to illustrate its features and to provide readers with an understanding of some of the design considerations in the system.

As mentioned, Family Portals (FP) was an always-on multi-family media space that connected three households together through a persistent video link. We called this media space "Family Portals" because it contained "portals," or views into distant families' homes (Figure 10.1). FP was prototyped on a touch and pen sensitive tablet PC with an external webcam to simulate the idea of it being a dedicated information appliance as opposed to a computer used for multiple tasks. The dedicated mobile device meant that FP could be easily moved throughout a family's home, depending on where household members wanted to share their activities from with a remote home. The user interface was divided between two types of portals: Targeted Portals for seeing individual homes and interacting with them on a one-to-one basis, and a Shared Portal for interacting with both families at the same time. We describe these next.

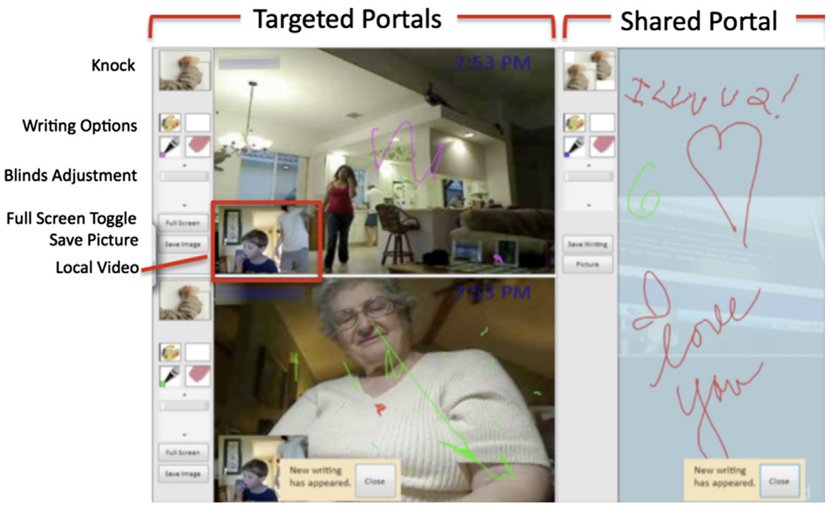


FIGURE 10.1

The Family Portals system showing shared video connections between three homes.

TARGETED PORTALS

The left side of the screen in Figure 10.1 shows two Targeted Portals (top and bottom), one for each family that a local user is connected to. The Targeted Portals were intended to allow families to interact and share information with *one* of the two remote families in a dyadic manner. The main portion of each Targeted Portal showed video from the remote family's home, which could be obscured, if the remote family desired, by adjusting "privacy blinds," as shown in Figure 10.2. The bottom left corner of each Targeted Portal showed a feedback view of what that particular remote home saw of the local user's home; again, blinds could be adjusted, depending on one's privacy needs.

FIGURE 10.2

Privacy blinds "pulled" halfway down at night in one home.

Within each portal, users had options for interacting with the remote family. An audio feed was not provided because of potential privacy concerns and the possibility of audio being too intrusive and distracting. Instead, families could write handwritten messages on top of the video feed, if they wanted to. They could also click a "knock" button to attract the attention of the family members at the remote location; the expectation was that family members would then approach their display.

We expected that families might want to focus their attention on one remote family at a time, for example, if the third family was not home. In this case, users could toggle between *Full Screen* and *Split Screen* views by clicking a button to the left of the portal. In Full Screen mode, the portal expanded to cover the entire screen, as shown in Figure 10.3. The third family's video was minimized and displayed at the bottom right corner of the screen.

FIGURE 10.3

A full screen view showing how a local home could focus on seeing one of the two remote homes.

SHARED PORTAL

The right side of the user interface was the Shared Portal (Figure 10.1, right). It provided shared interactions intended for the entire triad to see. Dourish et al. (1992) called this type of interaction *broadcast,* where all users of the system have access to all information. The main portion of the Shared Portal displayed a whiteboard, where users could write messages for *both* remote families to see. A multi-family knocking feature let a local family knock on both remote families' portals simultaneously.

THE EVALUATION OF FAMILY PORTALS

Following the design and implementation of FP, we wanted to learn how families would use the system in their daily lives, what communication and awareness practices would emerge, and what privacy concerns they would face. A field trial was the best method to explore these questions and to get ecologically valid usage data from real users who were using the system in their homes.

In the following section, we describe how we handled participant recruitment and deployment of the system, as well as the data collection and analysis. In the subsequent section, we talk about the issues and challenges that emerged during our study and how we overcame them.

PARTICIPANTS AND RECRUITMENT

When recruiting participants, we have found that a good approach is to have a multi-stage screening process. This may be time-consuming, but it ensures that the participants who are recruited match the needs of the study. For instance, in our study, we created an initial advertisement that broadly stated the requirements, timeframe, and compensation for the study and asked interested people to contact us by email. We posted this to local mailing lists (e.g., university students, internal company email lists, an email list of participants from previous studies we had conducted) and websites such as Facebook, Twitter, and Craigslist. When people contacted us expressing their interest, we sent them an email that asked screening questions about specifics such as what type of Internet connection they had and what their current practices were in relation to our research questions (e.g., how often do you use video chat?). We also asked potential participants if they had family or friends that they would like to connect with using a futuristic video system. (We were vague to avoid biasing respondents.)

This multi-stage recruiting process allowed us to generate a shortlist of potential participants for the study. We then followed up with people who met the demographics we were aiming for (e.g., families with children) to ask them if they and all members of their family were *interested and wanted* to participate. We found it especially important to ensure that all participating families were available to participate for the duration of the study and that none of the families were going be away from their home for an extended period. Since our study occurred during the summer months, when families tend to go on vacation with their kids, this was especially important to determine.

LESSON

Use a multi-stage recruiting process (e.g. advertisement, screener, and phone call) to narrow down a large respondent pool to a short list of suitable participants.

We selected six families (two triads) from the United States who we thought would provide interesting and diverse relationship dynamics. We believed it was important to pick participants who had a *real need* that could be fulfilled by using the system. For example, in this study, we recruited participants who were already communicating with remote households but wanted to *communicate more* and participants who were not communicating very often with remote households and wanted to change this pattern. If participants have a real need or can see the benefit of using a system, they will be more likely to adopt the system instead of just thinking of it as "something they have to use for the study." Tolmie and Crabtree (2008) found that conducting field trials with users who do not treat the prototype or system as an integral part of their home lives leads to data that may not accurately describe usage patterns or feedback about the system. Hence, determining participants' needs and sharing the value proposition of the system should lead to buy-in and higher user engagement with the system.

LESSON

Recruit participants who have a *real need* for the system and could benefit from using it as part of their everyday lives.

Our participant households were composed of young families, blended families, a divorcee, and retirees. We decided to select six families—two triads—because we thought that one triad would not be enough to understand usage patterns, as we would not have other families to compare the triad's experiences with. We thought that having three triads would be extremely difficult for a team of two researchers to manage (weekly visits with each home, ensuring that the system was constantly connecting all families within a triad, etc.). On the other hand, two triads seemed like a suitable number from which to gather rich qualitative data.

Next, we describe our participant families. Relationships are described from the point of view of the *seed family*, who responded to the advertisement for the study. A seed family is connected to two *remote families*.

Triad 1: The seed family (top of Figure 10.4) consisted of two parents and a three-year-old son. They used FP to connect with the wife's mother and stepfather and her maternal grandparents. In total, there were seven family members in this triad with diverse relationships and various communication needs. Figure 10.4 shows the families in Triad 1.

Triad 2: The seed family (top of Figure 10.5) consisted of two parents and a three-year-old son. They were connected to the wife's mother and her older sister and family. There were nine family members in this triad with diverse relationships and various communication needs. Figure 10.5 shows the families in Triad 2.

INITIAL INTERVIEW AND SYSTEM SETUP

The first stage of the study involved visiting each family's home and interviewing them about their existing communication practices with their extended family and, more specifically, about the families they were going to connect to using FP. One family lived outside our driving distance and was interviewed using Skype.

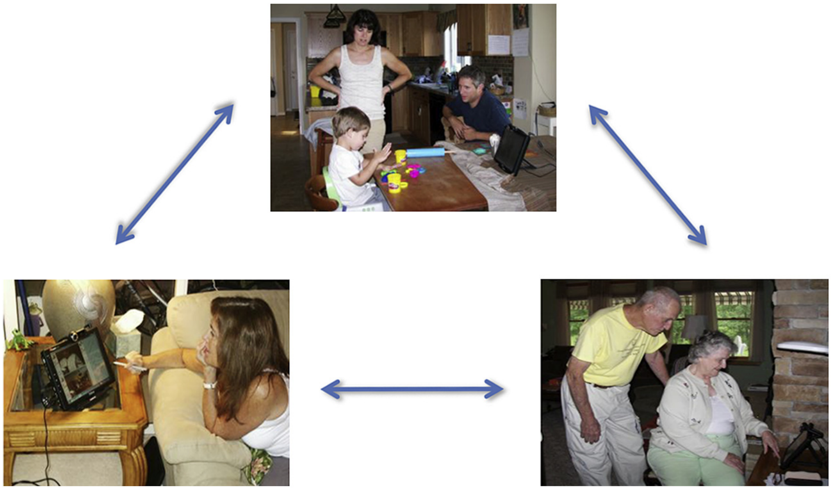


FIGURE 10.4

Participants in Triad 1 and location of Family Portals in their homes.

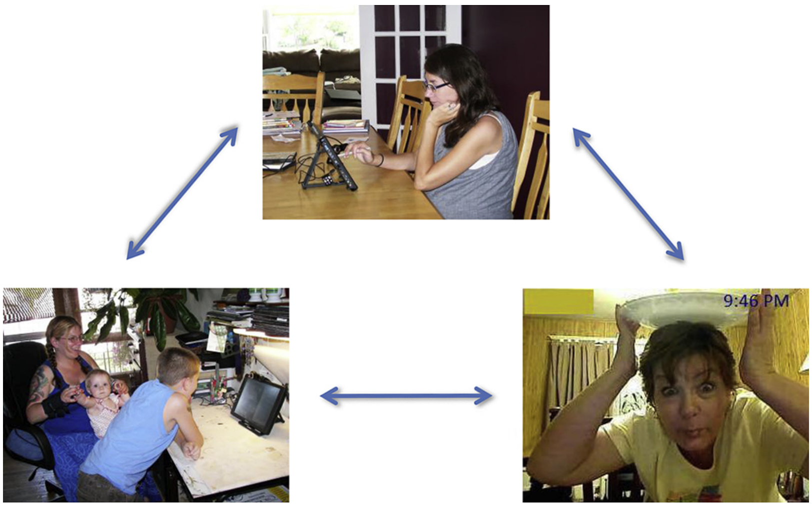


FIGURE 10.5

Participants in Triad 2 and location of Family Portals in their homes.

Chapter 2 provides more information about conducting interviews with remote participants.

We notified the local families beforehand that, since we were visiting on a week-night, when they might already be very busy, we would bring pizza and have a light dinner with the family. This helped us get to know the family in an informal manner before starting the interview. After the interview, families were walked through the initial setup of FP on their home wireless Internet connection and asked to select a location for FP. The initial interview lasted one to two hours, and setup took an additional hour.

Because of the pragmatics of the situation, we had only one researcher attend our initial interviews and setup of the prototype in participants' homes. Unfortunately, we learned the hard way that it would have been better to have two or three members of the research team present during the initial interview and setup. Since this is the initial face-to-face visit with a participant family, having more than one researcher present is essential in ensuring the researchers' safety in an unknown environment. And with at least two team members, one person can focus on interviewing and the other on note taking. Having a dedicated note-taker will allow the interviewer to focus on the conversation and develop rapport with the participants.

We encourage field teams to include at least one person who is familiar with the prototype or software during the initial home visit. This person should be able to debug and troubleshoot issues that come up during setup and installation. This will also allow the interviewer to focus on interviewing the users without having to worry about installation issues.

LESSON

Field teams should comprise two or three people with separate roles such as interviewer, note-taker, and software expert.

DEPLOYMENT AND INTERVIEWS

Getting the system up and running in participants' homes is a huge milestone in a field trial. However, the research team cannot sit back and relax once this occurs. We would argue that the actual deployment period can be the most challenging part of a multi-home field trial because one needs to ensure the system is running normally for all households and be available if a situation goes awry. After the initial interview and system setup, we found following up with participants via email or phone calls at predetermined intervals was critical to ensure that they were using the system and that the system was working as intended. We found that we frequently had to play the role of technical support; prototype systems are often frail and plagued with errors and issues, and FP was no exception.

All six families had FP in their homes for a period of eight weeks. We visited families throughout the deployment and conducted weekly, semi-structured, contextual interviews with them during this time. The one distant family was interviewed weekly using Skype. Questions focused on families' usage of FP, changes in communication, connectedness, and awareness of remote families, and privacy concerns. Adults were interviewed individually when possible, and children were interviewed with parents present. We interviewed adults in the family individually to learn about changes in the family dynamics and the effects of being connected through always-on video with not just one but two other families. Each family was given a diary to log self-reports between interviews. In addition to this, we sent emails and phoned participants between interviews to check for technical difficulties and to troubleshoot problems.

We removed FP from the families' homes at the end of eight weeks, after we conducted end-of-deployment interviews. We conducted post-deployment interviews two weeks later. Questions for the post-deployment interview focused on changes in communication, connectedness, and awareness of remote families' lives in the previous two weeks compared to the eight weeks when the families used FP in their homes.

Although FP was placed in families' homes for eight weeks, they were only able to use the system for an average of five weeks. Despite rigorous testing before deployment, numerous technical issues related to connectivity arose. Nonetheless, we found that five weeks of use was sufficient for participants to overcome novelty factors, develop patterns of usage, and incorporate the system into their daily lives. We discuss these technical challenges in more detail in the next section and share tips for overcoming them.

DATA COLLECTION AND ANALYSIS

Given that field trials are unpredictable, we have found it is very important to utilize multiple data collection methods. This is especially helpful if one method does not gather enough data. For example, solely relying on usage data from system logs may not be sufficient, as the system might be down for a few days, families may have issues with their Internet connection, or a family member might accidently drop and damage the system. Having multiple data collection methods will also lead to richer data and more engaging user stories, allow for comparing responses from different participants, and fill gaps in the data through triangulation (Creswell 2003).

We collected data using multiple methods, including interviews, diaries, observations, and client-side system logs. All interviews were audio recorded, and handwritten notes were taken to aid analysis. We also observed families interacting with FP either before or after some of our interviews. In total, we acquired data from approximately 108 hours of interviews and observations across all six families. The interviews were the primary sources of data, as participants would often forget to fill out the self-report diaries. This is a common problem in diary studies, which makes relying on diaries as a primary source of usage data difficult. Our system also logged the usage of features (e.g., blinds, full screen) throughout the study. Having multiple sources of data gave us the opportunity to gather as much information as possible during the field trial.

LESSON

Use multiple data collection methods to ensure there is enough data to analyze after the field trial.

During the analysis phase, we used open coding (Corbin and Strauss 2008) to analyze the interviews, observation notes, and diary entries. We generated codes that reflected a variety of usage patterns, which are described in their entirety in Judge et al. (2011) and Judge (2011).

In the following sections, we describe the challenges we faced and lessons we learned through our field trials of FP.

PROTOTYPE FRAILTY AND PILOT TESTING

Perhaps the largest challenge we faced in our study of FP related to the frailty of our prototype. Field trials are typically conducted with research prototypes instead of high-quality finished products. This is because the design is still being refined and has not gone through a commercialization process. A prototype is by definition unfinished, which means it can be buggy and prone to having technical issues. FP was no exception, despite rigorous testing prior to the study.

We designed and developed FP in a work environment at Kodak Research Labs. This meant that most of the initial testing of FP was done in the lab on a local intranet (internal to Kodak) with a bandwidth higher than what is achievable in users' homes.

Given that most prototypes are conceptualized and created in a lab environment, it is easiest to test the system in the same lab environment. However, a lab environment setup is very different from users' homes, and it is easy to be unaware of technical challenges that may occur in the field without conducting pilot tests away from the lab. For instance, most academic and industry labs have high-speed Internet access and a firewall. This setup does not necessarily translate to those in the homes of users. Homes may easily have lower Internet bandwidth that can vary depending on the time of day or number of devices utilizing it. Family members are responsible for ensuring their Internet connection and devices within the home are working, which contrasts with work settings, in which there are often trained technical staff who monitor and fix issues related to Internet connections, computer software and hardware. Because of this, we decided to pilot test FP outside of the lab to find additional problems and bugs that we could fix before the field study. However, this raised several issues.

FAMILY PORTALS SYSTEM ARCHITECTURE

First, we had issues related to the underlying system architecture we used for the system. FP had been designed to use a client-server architecture. This meant we launched specially-designed server software on a computer and each FP system acted as a client and connected to it. The server software then acted as a central distribution point for data sent between each client. For example, video from any client would be sent from the client to the server. Upon receipt, the server software would then retransmit the video data to the other clients. There are certainly other ways to design a networked system like FP, for example, by using a peer-to-peer architecture. Yet we wanted to ensure that any of the three households using the system could have the ability to turn their video on or off such that they could join or leave the media space on their own accord. This would create a "persistent place" that any client could connect to. In our view, a client-server architecture lent itself most naturally to this situation.

However, this created problems as soon as we tried to pilot test FP outside of the lab prior to the field trials. We ran our server software on a server computer connected to Kodak's internal network, which had an external IP address that one could use to see it from the "outside world." Yet it was still behind the Kodak firewall and, as such, any clients running from outside Kodak's internal network could not reach it to connect to it. Given that we were testing a multi-home system, all three instances of the system had to connect to the same server to establish a persistent connection. Because Kodak had privacy and security concerns about corporate data, it was not possible for us to create the ideal server setup within a reasonable timeframe. To circumvent this, we moved our server to the home of one of the researchers to be uninhibited from networking restrictions typical with most companies. This worked reasonably well, though it did mean that we were restricted in terms of Internet bandwidth, which is typically less for home Internet connections compared to corporate connections.

SIMULATING A THREE-WAY CONNECTION FOR TESTING

Second, we found it difficult to simulate a three-location connection for pilot tests prior to the study. We wanted to test the system in a setup that was similar to the real setup we would use in the field study: three homes connecting together, with our server software running in a fourth location. Connecting one location to the server was easy because we could have one project team member run a client on a home network and another run the server on his home network. However, moving beyond this to a true multi-family situation was challenging. We tried testing with three employees, each at their own home; however, this involved recruiting additional colleagues beyond our core design team, and organizing and coordinating became challenging.

In the end, we were not able to fully simulate the real setup that we were going to use in our field trials. Yet we believed that we had uncovered many of the issues we would face in our study and that our pilot testing was good enough to ensure few issues would occur during our field trials. We were, unfortunately, wrong, and this meant we spent many painstaking hours during the first three weeks of the field trial fixing problems and issues that arose based on the real-world setup of the system in the families' homes. The lesson is that testing a system in a real situation is extremely valuable; however, the reality is that, with multiple connected households, adequately carrying out such pilot testing can be challenging.

> **LESSON**
>
> Test and retest the system setup and installation in environments outside of the research lab that are similar to the real-world setting in which the field trial will take place.

INSTALLING THE SYSTEM IN MULTIPLE CONNECTED HOUSEHOLDS

After our pilot testing was completed, we conducted our initial visits with families and set up FP in their homes. This revealed several interesting challenges (and lessons) that increased the length of our initial home visits beyond our initial expectations of a couple of hours. These challenges also created more technical hurdles than we expected.

HIGH-SPEED AND RELIABLE INTERNET CONNECTIONS

We quickly realized that FP would not be an isolated system during its testing. It would become part of a larger ecosystem of devices and infrastructure in participants' homes. As researchers, we might think that we are installing a new piece of technology that will be separate from other devices in the home. However, the reality is that new technology often has to fit within an existing ecosystem of devices and technology set up in the home. In our case, it was of utmost importance to get the prototype connected to wireless routers and Internet connections in participants' homes. We had to play the role of technical support specialists to get the prototype up and running while troubleshooting network connectivity issues.

For a multi-home deployment, we found it was important that all homes had an Internet connection with high bandwidth, because if the Internet connection was slow at one home, it affected the experience at the other homes. This is quite different from a single-home deployment, where, if the Internet connection is down or slow in one home, it does not affect any other participants in the study. In single-home deployments, the system can be programmed to work in an offline mode, which synchronizes data with a server once the Internet connection is back up. In our case, such a setup was not possible because video had to transmit in real time between home locations. We had naïvely forgotten to ask about Internet bandwidth in our screening questionnaire. And often, the actual home Internet bandwidth is lower that the bandwidth that the provider advertises.

In retrospect, we also found it was important to know, prior to our initial home visit, whether participants' homes had a wireless router, which would allow participants to place FP in a location of their choosing. Ideally, we did not want the device to be restricted to one location or tied to the nearest wired Ethernet port. At the time of the study, wireless routers were commonplace, and we had anticipated that all of our participants' homes would naturally have it. Again we learned that our assumptions were wrong, and, unfortunately, we had not included this question in our screening questionnaire. In one instance, we drove an hour and a half on a Sunday evening to get to a participant's home to set up FP, only to find that they did not have a wireless router. We then had no choice but to drive around an unfamiliar town at 7:30 pm looking for an open store where we could buy a wireless router for the participant's home.

LESSON

Before the field visit, identify the equipment and technological setup that the participants' homes need to have. Do not assume that all homes have high bandwidth Internet access and other pertinent hardware (e.g., wireless routers).

We learned that, even if participants have the necessary hardware such as a wireless router, they need to know any setup information for it or have this information available during the initial study visit. In our case, several participants did not remember their wireless router password, and we did not ask them to find this information before we arrived. This caused delays in setting up the FP prototype while family members searched for their instruction manuals and tried out various passwords. It also meant that, in some instances, we had to reset their router so that the password would reset as well. If this happens, the researcher might also be asked to help the family reconnect their existing devices to the wireless network, as we were. Be prepared to receive frantic phone calls when, later, they try to connect other devices to the home Internet connection and are unable to do so.

From our experience conducting field trials, we knew that each home typically has a tech-savvy family member who deals with things like connecting devices to the Internet. It is a good idea to identify this person ahead of time and have him or her present during the home visit to aid in connecting devices and altering a family's home Internet infrastructure.

While the above challenges mostly relate to Internet connections and wireless routers, which will naturally improve over time and become standard devices in most homes, the lessons certainly apply to any type of hardware or connectivity that a prototype system relies on. In our case, it was a wireless router, and in other cases, it may be a new piece of hardware of the future that some families have and others do not.

SETTING UP THE PROTOTYPE IN A REMOTE LOCATION

In our study, we worked with families who lived in remote locations. Given that we were studying how families connected with each other over distance, it was natural that our triads were spread across multiple states within the United States. This made it impractical to travel to each location to set up the prototype in person. We were lucky in that only one participant family lived farther away than driving distance; we were located in Rochester, New York, and they were located in Florida, about 1,500 miles away. As such, we mailed them the prototype system in advance and scheduled a time to use Skype to walk them through the setup. Fortunately, the prototype was not damaged during the shipment, which is certainly a concern when shipping computer hardware.

When we conducted our virtual initial visit with the participant family in Florida, we experienced similar setup issues as we had with the families we had set up in person. This related to being unable to connect the FP's tablet to the home's Internet connection. The problem was the same as in other homes, yet now we needed to manage it over a Skype connection and have the family members solve the problem, as opposed to us solving it. In this case, using a Skype connection was extremely helpful because we could ask the participants to turn the FP tablet to face their laptop webcam so we could read error messages that appeared and guide them. It was also helpful to have a similar tablet PC device next to us that we could look at to determine troubleshooting options. Eventually, we were able to successfully set up this remote family with FP; however, we learned that walking remote families through the installation of our prototype took at least twice as much time as it did with a local family. This has implications for scheduling the initial home visit to ensure there is enough time to adequately work with the family to set up the prototype. It also means letting the family know that the initial visit may take a substantial time commitment on their part.

LESSON

Installing and setting up a prototype in a remote home takes longer than setting it up in person in local homes.

CONNECTING ACROSS LOCATIONS

We found that, because we were connecting across multiple locations, our installation and setup of FP was iterative. That is, connections would happen in stages as we set up each family. The first family who was set up would not be able to see anybody else on the system until we set up the second family. Once the second family was set up, they would only see the first family. It would not be until the third family was set up that all three families could see each other. Overall, this meant that we had to check in with families multiple times after their initial setup procedure to ensure they could see their remote families, who were being iteratively set up with the system.

To avoid issues with this iterative setup process, we tried to be strategic, such that we made sure the most tech-savvy family was the first that was connected. This way they could help other families when they were getting set up and connected. This approach also made it much easier for us to explain the functionality of the prototype to the second and third families, because they could see live video feeds coming from the other homes. We ensured that one or both of the remote families were available when we installed the prototype in a local family's home. By doing this, we gave the participants an opportunity to learn about and test the functionality of the prototype in our presence. Added benefits of doing this were that we were able to observe the initial interaction between families and note initial impressions and feedback. This was valuable data, as it gave us a baseline to compare against in the following weeks of usage.

LESSON

Consider setting up the system in the home of the most tech-savvy family, followed by the less tech-savvy families. This will help the less tech-savvy families easily understand the system when it is installed in their homes.

OWNERSHIP AND USAGE OF THE SYSTEM

We recognized in our field trials that it was very important to ensure that participants felt a sense of ownership over the prototype that we were giving them. We found that this was an important factor in the adoption and usage of the prototype and also led to rich data from our participants. Giving participants ownership will prevent them from just thinking of the system as "stuff" in their home that has low value (Tolmie and Crabtree 2008).

We encouraged participants to feel a sense of ownership of the system in two ways that worked out well. This was evident by the way participants talked about the system as being theirs, as well as how they moved it around in the home on their own accord and decided how they would use it.

PICKING OTHER FAMILIES TO CONNECT TO

During our recruitment, we had asked our seed families to pick two other families they wanted to connect to using the system. By doing so, we gave them ownership over the system and its use and provided them with a sense of agency as part of the study process. That is to say, they themselves were in charge of whom they would use the system with and how they would use it. We relied on the seed families to socialize the idea of FP to their extended families as part of the recruitment process, rather than us telling the other families about the system and its features. By allowing seed families to pick other families they wanted to connect to, we learned about their current communication practices with the other families and how they would like to use the system to improve or change their communication practices and, at times, even their relationships. Overall, this meant that, after we identified seed families, we were able to easily get an additional two families in each triad to participate in the field trial.

DECIDING ON THE PLACEMENT OF THE SYSTEM

We gave participants ownership over the system by allowing them to choose where to place it in their homes. Ownership, as it relates to device placement, manifested itself in two ways with the families. First, we found that some families had a default placement for the system that did not change much from the initial placement of the device. The device was then moved around the home based on where family members were gathered or where they were doing activities that they wanted to share with the remote families (e.g., kitchen, porch). For instance, Figure 10.6 shows that a remote family temporarily placed their FP on their porch to share their garden, and another family temporarily placed FP in their family room to share a child's play time with remote families.

Second, we found that some families frequently moved FP for the first week until they found a placement they were comfortable with. After this point, these families did not move the FP around their homes, and the remote families saw only one constant view. The primary reason for doing this was privacy. For instance, one of the sisters in Triad 2 was not comfortable sharing all of the activities in her home with her older sister's family and her mother, so she placed the FP on her dining table facing a wall and only turned it around to face the family room when she was communicating with the remote families (Figure 10.7). This led to many complaints from the remote families that they were only seeing her wall.

Together, these findings illustrate the importance of supporting a diversity of placement options such that most families will be able to gain a sense of ownership based on device placement and usage.

FIGURE 10.6

Family Portals was moved around the home based on where family members were gathered. On the left, the remote family placed their Family Portals on their porch, and on the right, the local family placed Family Portals in the family room to share a child's play time.

FIGURE 10.7

Family Portals in the younger sister's home was placed on the dining table facing a wall.

LESSON

Give participants ownership over the system by allowing them to pick whom they use it with and where it is placed in their homes.

MANAGING THE COMPLEXITY OF RELATIONSHIP DYNAMICS WITHIN AND BETWEEN FAMILIES

Having discussed technical challenges and lessons learned from deploying a multi-home system, we now discuss another challenge in multi-home field deployments: managing and navigating the relationship dynamics within and between families.

VARYING LEVELS OF COMMUNICATION NEEDS WITH EXTENDED FAMILY

Although all families agreed to participate in our study, there were varying degrees of need to communicate with remote families and varying levels of buy-in from family members. The primary, or most active, users of FP in each triad were the women. In Triad 1, this was the daughter, mother, and grandmother (Figure 10.4). In Triad 2, it was the younger sister, older sister, and their mother (Figure 10.5). This is consistent with the findings from Hindus et al. (2001), which show women are generally the "household communicators." It is important to note that one woman in each triad answered the advertisement for the study and determined who they would like to be connected to using FP. It is no surprise that they all chose to connect to their own families instead of their in-laws.

As we expected, there were certain individuals who did not like or were apathetic to always-on video being broadcast from their home. These users did not necessarily voice their concerns or their lack of need for the system at the beginning of the study, but their lack of use and, sometimes, even disdain for the system became more apparent as the study progressed. Findings from our initial field study of the Family Window showed that privacy concerns and lack of interest in the system were not issues in dyads where families shared a close relationship (Judge et al. 2010). However, this was not the case in triads, as there are more families and more individuals per triad. For instance, we found that the husband and the live-in boyfriend in Triad 2 avoided FP because they did not want to be seen by their in-laws all the time.

We also found that the grandfather in Triad 1 did not use FP because he was intimidated by the technology. After technical problems with the system (because of set up issues), he was afraid to use FP because he was concerned he "might break it." He did, however, occasionally look over his wife's shoulder while she interacted with the remote families.

LESSON

Not all family members are going to want to use a prototype or system that is placed in their homes. Be cognizant about this during the study and try to learn about the reasons behind the lack of usage.

The challenge here was developing rapport with the participants so they were comfortable enough to share such private and, at times, even embarrassing information (e.g., not wanting to communicate with his mother-in-law) with the research team, who were essentially strangers.

DEVELOPING RAPPORT AND TRUST WITH PARTICIPANTS TO LEARN ABOUT RELATIONSHIPS DYNAMICS AND USAGE PATTERNS

As we expected, learning about the primary users' successful adoption of FP was easy enough, but our findings would have been incomplete without learning about privacy concerns and other reasons for lack of adoption. We found that the way around this issue was to establish trust and rapport with the participants early on in the study. We did this by conducting separate interviews for each adult in a family and by repeatedly assuring them that the information they shared would not be shared with other families in the triad.

For instance, the older sister in Triad 2 shared with us that she was not happy that her mother was in a relationship with someone she did not approve of. Also, she quickly discovered that when her mother's boyfriend was visiting, her mother would turn off the FP camera or turn it to a different direction. Here, it was important for us to be mindful that this was private information and could potentially lead to conflict in the family if it were shared with the mother. However, it was important for us to learn more about this phenomenon from the mother's point of view.

To do this, we casually mentioned to the mother that her eldest daughter said there was no video feed from her mother's home a few times in the previous week. This led to the mother telling us about her boyfriend's visits, her daughter not approving of him, and hence, her need for some privacy from the always-on video. By approaching the issue this way, we *triangulated* (Creswell 2003) the data given to us by the eldest daughter with responses from her mother. Knowing both sides of the story enabled us to create a rich narrative that described the occurrence. We found triangulation to be an important interviewing and data analysis strategy to learn about FP usage. We also learned that ensuring privacy and confidentiality was important, especially when personal and often contradictory information was shared about other family members.

Another way to create rapport with the participants and make them comfortable during interviews was to have the same interviewer throughout the study. By doing so, we thought that participants would not have to start over with a different interviewer in each session.

LESSON

Developing rapport and trust with participants is important in order to learn about their usage and potential concerns with the system. It is also ethical to protect the privacy and confidentiality of the information shared by participants and to avoid sharing it with other family members.

CONDUCTING INTERVIEWS INDIVIDUALLY

When possible, we conducted individual interviews with the adults in each family. It was the only way we could find out what participants *really thought* about the system and about connecting with remote families via always-on video. For instance, these interviews allowed us to learn how a son-in-law felt about being connected to his mother-in-law and sister-in-law's home through the always-on video system. It would have been difficult to learn about his privacy concerns and discomfort if he had been interviewed with his wife, who was happy about being connected to her mother and sister.

However, we quickly learned that individual interviews were not easy to do because of families' routines and work schedules. For instance, the live-in boyfriend of one of the sisters in Triad 2 worked night shifts and was often asleep during the day when we visited their home. Because of the difference in work schedules between adults in the families, we changed weekly interview times so that we could alternate interviewing the two adults in the family. This was not ideal, but it was the only option for some families. In other families, especially families with young children, we interviewed one parent while the other was looking after the children in a different room and then had them switch places. This strategy, when it was possible, worked very well, as we were able to interview one adult without interruptions from the children.

LESSON

When possible, interview participants individually to learn about their usage of the system and potential privacy concerns.

We also found that we had to be flexible and open to interruptions during the interviews especially if there was only one adult present in the home with children. For instance, the eldest sister in Triad 2 had a one-year-old daughter, who frequently interrupted our interviews by crying. Sometimes this participant had to be interviewed in the kitchen while she was preparing dinner and helping her kids with their homework. The lesson here is to be flexible and respectful of things that might be going on the participants' homes, and be prepared for multiple interruptions and short attention spans from participants. These are some of the challenges of working with families, which are also reflected in other chapters in this book.

BEING RESPECTFUL AND REFLECTIVE OF YOUR ROLE IN PARTICIPANTS' LIVES

Because our study lasted eight weeks and we visited families weekly, we became privy to their daily routines as well as their ups and downs. For instance, one of the families in Triad 1 started treating the interviewer, a young graduate student at that time, as one of their grandchildren and would spend half the interview talking about their younger days and sharing health concerns and information about doctor visits. Being mindful of this is important, as participants may start treating members of the research team as part of their family, especially for longitudinal studies. This phenomenon is common in ethnographic studies, where a researcher becomes a part of the community he or she is studying (Spradley 1980). When this happened, we had to stay unbiased and not be emotionally attached to the participants. This was easier said than done, because we felt very welcome in some participants' homes, and we were even invited to their family get-togethers. As only one member of our research team was in the field interviewing users and collecting data, one of our methods to overcome potential bias in the data collection was to do *peer debriefing* (Creswell 2003). After each interview, the interviewer debriefed with members of the research team to share findings, identify potential holes in the data, and overcome any biases in the data analysis.

LESSON

Be aware and respectful of your role in participants' lives especially for a longitudinal study. Take necessary measures such as *peer debriefing* to ensure there is no bias in the data collection and analysis.

DISCUSSION AND CONCLUSIONS

This chapter describes the use of multi-home field trials as a method to evaluate and understand the usage of domestic technologies. We used Family Portals (Judge et al. 2011), an always-on video media space that connected three homes, as a case study to illustrate lessons and challenges in conducting a multi-home field trial. Although a multi-home field trial seems similar to a single-home field trial, this chapter highlights two primary differences between multi-home and single-home field trials: an increase in technological challenges and an increase in the diversity of relationships between participants in the study.

First, we found that a multi-home field trial is more complex and technologically challenging to set up compared to a single-family field trial. Since most prototypes are developed in a lab setting, it is important to pilot test them in a setting that is similar to participants' homes before the field trial. However, we found that accurately simulating a three-home connection with a server that was behind a firewall was difficult. The lesson is that pilot testing a system in the real situation is extremely valuable; however, in practice, adequately carrying out such pilot testing with multiple connected households can be challenging. We also found it important to know ahead of time about the reliability of participants' home Internet connections and to be prepared to play a technical support role during the study. An important lesson here is realizing that one is not installing a standalone system or prototype in a participant's home. Instead, the system is a part of a larger ecosystem of devices and technological setup.

Second, we found that there are many diverse and dynamic relationships within and between homes in a multi-home field trial, and researchers must be aware of these dynamics. One lesson is to be aware of potential conflict or discomfort that may arise from the longitudinal use of a system. To learn about this, interview family members individually and triangulate data. We found that developing rapport and trust with users early on in the study is important. This will help users feel comfortable to share information about their lives, which may be affected by using the system. In addition, as researchers, it is ethical to protect the information that is shared with us and to make sure information shared by one family is not revealed to another family, as this could potentially lead to conflict between families.

In conclusion, although multi-home field trials are challenging to set up and conduct, we believe that, if one is mindful of the lessons presented in this chapter, these types of field trials can be successfully conducted to evaluate the use of domestic systems. Although our field trial focused on connecting three households, we anticipate that the lessons we uncovered are applicable to studies of systems that include even more than three connected homes. Naturally, the complexity would increase beyond this as the number of homes increases.

ACKNOWLEDGMENTS

This research was generously funded by Kodak Research Labs. We thank our collaborators at Kodak Research Labs and Virginia Tech for valuable feedback on this work. They include, but are not limited to, Andrew Blose, Rodney Miller, Andrew Kurtz, Elena Fedorovskaya, Steve

Harrison, Manuel Perez, Dennis Kafura, Deborah Tatar, and Andrea Kavanaugh. We also thank all the families involved in the field trial for welcoming us into their homes and sharing their experiences with Family Portals.

REFERENCES

Barkhuus, L., Rode, J.A., 2007. From Mice to Men — 24 Years of Evaluation in CHI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York, NY.

Bly, S., Harrison, S., Irwin, S., 1993. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. Commun. ACM Press, 36 (1), 28—46.

Brown, B., Reeves, S., Sherwood, S., 2011. Into the Wild: Challenges and Opportunities for Field Trial Methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York, NY.

Brush, A.J., Inkpen, K., Tee, K., 2008. SPARCS: Exploring Sharing Suggestions to Enhance Family Connectedness. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. ACM Press, New York, NY.

Conversy, S., Mackay, W., Beaudouin-Lafon, M., Roussel, N., 2003. Videoprobe: Sharing Pictures of Everyday Life. In: Proceedings of the 15th French-speaking conference on human-computer interaction on 15eme Conference Francophone sur l'Interaction Homme-Machine (IHM). ACM Press, New York, NY.

Corbin, J., Strauss, A. (Eds.), 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage.

Creswell, J., 2003. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications.

Dourish, P., Bly, S., 1992. Portholes: Supporting Awareness in a Distributed Work Group. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY.

Froehlich, J., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, E., Fu, F., Bai, M., Patel, S., Landay, J.A., 2012. The Design and Evaluation of Prototype Eco-Feedback Displays for Fixture-Level Water Usage Data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY.

Greenberg, S., Buxton, B., 2008. Usability Evaluation Considered Harmful (Some of the Time). Proceedings of the SIGCHI Conference on Human Factors in Computing System (CHI 2008). ACM Press, New York, NY, pp. 111—120.

Harper, R.H.R., Lamming, M.G., Newman, W.M., 1992. Locating Systems at Work: Implications for the Development of Active Badge Applications. Interact. Comput. 4 (3), 343—363.

Hindus, D., Mainwaring, S., Leduc, N., Hagstr, A., Bayley, O., 2001. Casablanca: Designing Social Communication Devices for the Home. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2001). ACM Press, New York, NY.

Intille, S., Larson, K., Tapia, E., Beaudin, J., Kaushik, P., Nawyn, J., Rockinson, R., 2006. Using a Live-In Laboratory for Ubiquitous Computing Research. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A. (Eds.), Pervasive Computing LNCS 3968. Springer, Heidelberg, pp. 349—365.

Judge, T.K., 2011. Patterns of Domestic Video Mediated Communication. Doctoral Dissertation. Department of Computer Science, Virginia Tech, Blacksburg, Virginia.

Judge, T.K., Neustaedter, C., 2010. Sharing Conversation and Sharing Life: Video Conferencing in the Home. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010). ACM Press, New York, NY.

Judge, T.K., Neustaedter, C., Kurtz, A., 2010. The Family Window: The Design and Evaluation of a Domestic Media Space. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010). ACM Press, New York, NY.

Judge, T.K., Neustaedter, C., Harrison, S., Blose, A., 2011. Family Portals: Connecting Families through a Multifamily Media Space. In: Proceedings of the CHI 2011. ACM Press, New York, NY.

Kidd, C., Orr, R., Abowd, G., Atkeson, C., Essa, I., MacIntyre, B., Mynatt, E., Starner, T., 1999. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In: Streitz, N.A., Hartkopf, V. (Eds.), CoBuild 1999 LNCS 1670. Springer, Heidelberg, pp. 191–198.

Kirk, D.S., Izadi, S., Sellen, A., Taylor, S., Banks, R., Hilliges, O., 2010. Opening up the Family Archive. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW 2010). ACM Press, New York, NY.

Korn, M., Bodker, S., 2012. Looking Ahead: How Field Trials Can Work in Iterative and Exploratory Design of Ubicomp Systems. In: Proceedings of the ACM Conference on Ubiquitous Computing. ACM Press, New York, NY.

Lindley, S., Harper, R., Sellen, A., 2010. Designing a Technological Playground: A Field Study of the Emergence of Play in Household Messaging. In: Proceedings of the 2010 SIGCHI conference on Human Factors in computing systems (CHI 2010). ACM Press, New York, NY.

Mynatt, E., Rowan, J., Craighill, S., Jacobs, A., 2001. Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2001). ACM Press, New York, NY.

Olson, J.S., Teasley, S., 1996. Groupware in the Wild: Lessons Learned from a Year of Virtual Collocation. In: Proceedings of the ACM conference on Computer supported cooperative work (CSCW 1996). ACM Press, New York, NY.

Neustaedter, C., Brush, A.J., 2006. 'LINC-ing' the Family: The Participatory Design of an Inkable Family Calendar. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006). ACM Press, New York NY.

Neustaedter, C., Brush, A.J., Greenberg, S., 2007. A Digital Family Calendar in the Home: Lessons from Field Trials of LINC. In: Proceedings of Graphics Interface 2007 (GI 2007). Canada, Montreal, Quebec.

Nielsen, J., 2004. Risks of Quantitative Studies. Nielsen Norman Group [online]. http://www.nngroup.com/articles/risks-of-quantitative-studies/.

Plaisant, C., Clamage, A., Hutchinson, H.B., Bederson, B.B., Druin, A., 2006. Shared Family Calendars: Promoting Symmetry and Accessibility. ACM Transactions on Computer-Human Interaction (TOCHI) 13 (3), 313–346.

Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R.E., Hursey, J., Toscos, T., 2007. Why It's Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp. In: Proceedings of the Conference on Ubiquitous Computing. Springer-Verlag, Berlin, Heidelberg.

Rowan, J., Mynatt, E.D., 2005. Digital Family Portrait Field Trial: Support for Aging in Place. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005). ACM Press, New York, NY.

Spradley, J., 1980. Participant Observation. Wadsworth Publishing Company, Belmont, CA.

Tolmie, P., Crabtree, A., 2008. Deploying Research Technology in the Home. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW 2008). ACM Press, New York, NY.