Interview and Observation for Mobile App Development

John Rooksby
University of Glasgow
John.Rooksby@Glasgow.ac.uk

MobileHCI 2013, Aug 27 – 30, 2013, Munich, Germany.
ACM XXX-X-XXXX-XXXX-X/XX/XX.

Abstract
This position paper discusses how user feedback, particularly from interviews and observations, can be used to inform mobile app development. I argue that it is feasible to "scale up" such methods in order to help designers gain insight into large-scale deployments. The paper points to some technical approaches that may help in achieving this. The paper ends by arguing that scaling up field methods branches them away from ethnography. Ethnography has traditionally discussed scale in orthogonal ways to computer science.

Author Keywords
Ethnography; Large Scale; App Design;

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Using field methods such as interview and observation is often seen to be good practice in designing. By taking the time to communicate with people and observe them using technology, designers can get insight into what they want, need and do. Interview and observation enable relatively open exploration, they leave room for surprise: not everyone (and

sometimes not anyone) will use or experience technologies as the designers imagine [1][2].

One of the perennial problems of field methods is scale. How can designers engage with users when there are tens of thousands of them distributed around the world, using the app in short bursts, and often just once or twice before giving up? McMillan et al [3][4] have argued that it is possible to conduct interviews and (indirect) observations in such situations. They use examples of their engagements with players of an iPhone game. However, an alternative, quantitative paradigm is emerging in app design that attempts to understand use through statistical analysis of log data. Logging is routinely built into mobile apps, and advances are being made in machine learning that stand to have a significant impact on how apps are developed (e.g. [5]). So the question begs, if you have a log of what every user has done with an app and are able to draw inferences from it about app use, what forms of qualitative research are worth the effort? Even in relatively small deployments, fieldwork is sometimes said to be "not worth the hassle" [6]. The work reported by McMillan et al [3][4] was clearly labor intensive. In this paper I will consider how and why field methods remain relevant to app design, even where extensive logging and machine learning is employed and used to inform re-design.

Interview and Observation

There are many ways to do fieldwork. Following McMillan et al [3][4], this paper focuses on interview and observation (of course there are many other approaches to fieldwork and, indeed, there is much more to fieldwork that selecting and using a method).

Interviews, in a design context, can be conducted with users or potential users of a technology. Most interviews conducted in HCI are "semi-structured" in that the interviewer has a number of questions they would like to ask, but is open to directing the interview along whatever paths emerge as it progresses. Interviews are relatively straightforward to conduct - as long as the interviewer is able to stay quiet long enough for the interviewee to speak, they will usually turn up interesting information. The more problematic aspects of interviews tend to be in recruiting people and scheduling the meeting. Making something of the information afterwards can also be hard work. Transcribing interviews can be particularly time consuming. McMillan et al [3] report that interviewing people that have downloaded an app through an app store is less straightforward than in local trials. Users and designers do not have the same relationship in these contexts, and unless local players are recruited it is not possible to meet face-to-face. McMillan et al used log data to select interviewees and communicated with them in-game. They used responses to inform a further design iteration.

A problem with interviews is that they solicit post-hoc accounts of what people do. Such accounts will gloss over or ignore many details of what the interviewee does with a technology, and so may not cover the detail a designer is likely to be interested in. In one sense a "good interviewee" can orient to what the designer is interested in, however the danger is the interviewee focuses overly on what (they believe) the designer wants to hear [7]. Observations offer an alternative. Particularly where video is taken, the observer will be able to scrutinize in detail how a technology is used. As with interviews, observation

and video requires the person doing it to shut up and allow the user to get on with things (which can be frustrating, particularly when watching someone struggle with something you have designed). There are several ways in which observations can be made, perhaps in person, perhaps with a video camera or perhaps by installing a screen recorder on someone's device. It can be helpful to combine these, for example a video camera can capture someone's gestures and body movements but may not capture what is on screen so well. A screen recorder captures the screen well but just the screen. In practice, working with and playing back multiple video streams can be cumbersome. Some observational studies will be researcher provoked, in that people will be asked to perform tasks in front of a camera, perhaps talking aloud as they go. Other studies will be more fly-on-the-wall. There is sometimes a concern that videoing someone changes their behavior. But this assumes that there is a natural or normal behavior that is out there to be captured, and it assumes that the effects of videoing will somehow render the study useless. Videoing does influence people's behavior, but usually in ways that are momentary and palpable rather than pervasive, and these are often of little consequence [8]. Even where something unusual happens, the observation can be interesting. Participants will sometimes announce "this doesn't usually happen" which will preface some explanation or account of ordinary action. As long as we are open to observations being just that, rather than preoccupied with finding "natural" action, these occasions can offer interesting insight. From this perspective, observational studies begin to overlap with interviews. As with interviewing, setting up and then analyzing an observational study can be far more time consuming than actually conducting it. Transcribing or coding a video can take a large amount of time. McMillan et al [3] discuss use of video in mobile app trials in terms of "direct" and "indirect" studies. Direct studies require the researcher to be present and observing the user, whereas indirect utilize screen recording. Direct observation enables wider coverage (for example embodied action), but indirect is more practical for gathering global data.

Interview, Observation and Logging

So why conduct interviews, observations and other forms of fieldwork if it is possible to systematically collect logs about app use? Although McMillan et al [3] used logging in their studies, they used these only to characterize and track their user base. New advances in machine learning allow for more log-driven approaches to understanding users.

Not everything is logged. Possibly the most important point to be made about logging is that they do not capture everything. Logging is itself designed, the developers have to make decisions about what to log and how. Logging is also embedded into software and is triggered by changes to software states – which is not necessarily representative of everything done.

Logs are easy to misinterpret. As an example of misinterpretation, the logs from a pedometer app developed at Glasgow University showed a team that appeared to using the pedometers regularly. However, this interpretation was wrong; in interview it turned out that they thought it was "a shaking game" and had kept it running only because they did not know how to turn it off.

Information may be hidden in the logs. Fieldwork may reveal or inspire ideas about factors that are findable within the logs but which have not been previously recognized as interesting. The mere fact that you have a log does not mean you are making the most of it.

The user experience is not readily reducible to a series of logged events. The user experience is more than the sum of logged actions. The user may have tried to do things that were not valid actions and therefore not logged. Things like reading, finding things difficult or intuitive, enjoying or disliking images, finding things funny, and so on are not readily findable from a log. The user experience also relates to factors beyond the use of the app, such as installing, uninstalling, and when moving from and to, or integrating with other services.

Logs may be erroneous. Sometimes fieldwork can be used to identify problems with logging, although my experience thus far is that people have great difficulty in saying whether logged information is accurate. Even where a diary is kept, the study participant may have been thinking about and describing what they did in different terms and with casual attention to time. This kind of verification appears best used when the data already looks erroneous.

Collection of opinions, ideas and contextual information. Participants may have opinion and ideas that are helpful to the ongoing design. These may not always be good or correct, but are worth listening to. My experience is that it can be helpful when a participant mentions something you have thought of previously but not prioritized. User (mis)interpretations of what an app is, or is for, can also provide inspiration. Participants will be able to give contextual information about their app use. For example they may talk about or refer to what other apps they use, or perhaps if there are factors in their life that are influencing their behavior (e.g. it is exam time for students).

Interview, Observation and Designing

So how can interview and observation fit with design? What do we know about designing? Designing tends to be improvised and loosely structured around the IDEs and other tools available to developers rather than following a defined procedure. That designing does not follow a strict plan or method does not mean it is slap-dash or unprincipled. Designing, as characterized by Schön [9], Sharrock [10] and others, is methodical whether or not a process is adhered to. Some common features of designing are:

Problems become clarified during the course of designing. Design tends to focus on a "gap" rather than a specific problem. For example "how can I make a novel and fun exercise game". The exact problem (or problems) that designing addresses will emerge as the work progresses.

Designing is iterative. Software products tend to be developed iteratively, rather than as a one off release. This may be to offer improvements, to respond to a changing context or a changing market place. Design practice is also highly iterative between releases. Designers can be seen to regularly go back to and reflect on just what they are trying to do and have achieved as they progress.

Ideas about users pervade design. Another feature of design, particularly product design, is the fluid way in

which ideas about users are played into the design process. Designers talk sometimes about specific users, but also regularly play in typifications of users, as well as general notions of what any person would do or want (including the what the designers themselves would do). Ideas about the user are not confined to specific periods of user consideration, but are brought into many technical discussions and decisions [10].

Designing requires compromise and tradeoff. Design needs to aim for (if not hit) deadlines, budgets and so on. This means designers can never do everything they wish and need to make tradeoffs.

Designing is rarely 'greenfield'. Design does not happen in a vacuum. In app development, designers do not work from scratch but develop for devices, and integrate with services that have their own constraints, limitations, conventions and rules.

In order for interview and observation to be viable outside academic studies of app use, I suggest they are treated as a means of collecting vignettes and stories about use that the developers can be exposed to. So this way the fieldwork is not transcribed and analysed but is something that the developers are exposed to (in one form or another). Studies should only be undertaken systematically when there is a specific issue identified as interesting or problematic. Studies may best be conducted a short while after a release. At this stage of design, the designers should be more open to ideas rather than in the midst of working on new ones. The vignettes and stories should not be used to produce specific requirements but to inform how users are then discussed in the course of further designing. Often fieldwork will not supply definite things to do or implement, but things to watch out for, avenues to explore – things that help articulate the "gap". Often fieldwork will have a reciprocal relationship with approaches to logging and machine learning. Quantitative approaches should inform how decisions are made about who to communicate with and what to look for, but qualitative approaches should inform how logs are made and analyzed.

I do not believe the suggestions made in the previous section require any form of detailed or professional analysis to achieve. They are the sorts of things that can arise within the course of discussion by most developers with a user. It is not necessary for developers to do sociological research, just to gain insight. Arguably it is not even necessary for the developers to write down anything from the interviews (let alone transcribe them), although one of the issues here is that it can be quite hard to remember everything that was said without a transcript and it can be helpful when presenting topics or themes from interviews to give examples of exactly what was said.

Support for Interview and Observation

From the perspective of this paper, the challenges and ambiguities associated with interviewing and observing are not so much in how to analyze data but are more pragmatic issues in getting the studies done. Along these lines I suggest:

Build dynamic recruitment mechanisms into an app. It would be helpful to be able to recruit participants for interview and observation (or survey) through the app itself. Being able to select and target users based on particular characteristics will be helpful. McMillan et al

[3] recommend in game rather than external avenues of contact.

Find ways of rewarding participants other than payment. Payment helps ensure a good response to requests for study participants. In a global market place, it can be hard to pay interviewees. Questions such as whether people will interview for things like extra points, in game credits, or kudos have only begun to be answered by McMillan et al and others.

Partially automate the interview process. One thing tried by developers at the University of Glasgow has been to incorporate a verbal feedback form into a game. This was used by many people, but rarely in the desired way. People sang songs, just said "hello", and so on. It was also very difficult to sift through the many audio files to find anything or sense. This does not mean this form of incorporated interview is unworkable, but perhaps that it should offer more structure and that consideration needs to be given to who is going to listen to the data and how, rather than just how it will be captured.

Enable video feedback. It is, in principle, possible for users to screen record their use of an app, and to talk aloud as they do so with the microphone on. This may be useful as a way of feeding back information about bugs or difficulties, but would be interesting as a way of gaining insight into ordinary app use. However, even though this is technically possible, careful design and consideration will need to be put into how users can be encouraged to do this and to communicate in a meaningful way. Attention will also need to be paid to how the videos can be sifted and viewed.

Recruit local groups. Finally, as suggested by Morrison et al [11], recruiting local groups for face-to-face interviews and observations may remain the most effective way of collecting rich information.

This list of suggestions is technology centric, and it is important to note that doing fieldwork effectively at scale is not just a technical problem. Gathering the data is one thing, but work with it effectively is another, and that is an organizational issue.

Discussion: Rethinking Fieldwork and Scale

In this paper I have presented a view of fieldwork that sees it as constituent to design, something that is done by designers and is heavily tied in with logs and log analysis. To this point in the paper I tried to avoid the word "ethnography". What I am envisioning branches from ethnography. I think it is important to acknowledge that ethnography approaches scale orthogonally to computer science.

It is hard to study ethnographically lots of people in lots of different places [12]. But this is not to say ethnography doesn't scale. Scale in ethnography is not the sum of the people studied. It should not be treated as if it focuses on the same phenomena as quantitative work, only in more depth. Ethnographies have traditionally not been of individuals, but of cultures. Culture is a problematic term, but however it is taken, a culture is not the sum of its people. It is more so the order and practices by which they live. Ethnographies have traditionally been global, often comparing between cultures. Ethnographies have also rarely been of just the present, but have looked to the background and history of practices and places in order to understand them. An examination of the HCI and

particularly the CSCW literature will show that ethnography has and continues to address scale, but perhaps in terms researchers might not always readily recognize. For example, ethnographers address:

Perspicuous settings: Many of the earlier ethnographies published in CSCW were of control rooms. By conducting "ethnographies of the center" [13] these studies gave insight into how large scale coordinative practices were orchestrated.

Multi-sited Ethnography: These are not simply studies repeated in several places, but recognize that culture is constituted by intersection and flow. These studies follow people, connections and associations [18].

Infrastructure: A body of ethnographic work has focused on the development of infrastructure. This work examines how we can design for and within "the long now" [14][15].

Biography of artifacts: Designing is almost never done from scratch. Technologies evolve and new technologies emerge from and among older ones. Biographical approaches focus on how historical decisions and trajectories influence the now [16].

Patterns: In recognition that people act and interact in orderly and methodical ways, some ethnographers have looked to produce more abstract representations of interaction. For example Martin [17] developed ethnographic pattern models.

Ethnography can address large-scale systems without having to actually "scale up". These studies can and do impact the way technologies are designed. However, such studies are usually carried out orthogonally to design work itself – these studies are unable to select and address the specific issues designers are interested within the time frame of a design project. Rather ethnographers tend to work separately, producing broad implications for design. This is valuable, and I believe it is important to see studies of app use as branching from mainstream or academic ethnography.

What I am suggesting in this paper is not ethnography (or is not traditional ethnography) but is fieldwork for app designers for use within specific design projects. This fieldwork is not intended for the production of deep insights – it is not intended for the production of academic papers whatsoever. Rather it is intended to expose designers to how their technology is used. The approach here attempts to scale-up fieldwork along quantitative lines. It takes a more quantitative view where scale is the sum of users and uses of an app. This approach will inevitably subordinate field methods to logging and quantitative methods. This may not be a bad thing, but should not be seen as a vision for ethnography itself.

References

[1] Ulrich, K. (2012). Design: Creation of Artefacts in Society. Pennsylvania: University of Pennsylvania.

[2] Randall, D., Rouncefield, M. (2007). Fieldwork for Design: Theory and Practice. New York: Springer.

[3] McMillan, D., Morrison, A., Brown, O., Hall, M., Chalmers M. (2010) Further into the Wild: Running Worldwide Trials of Mobile Systems. Pervasive 2010: 210-227.

[4] McMillan, D., Chalmers, M. (2012) Ethnography for Large Scale User Trials. Proceedings of the 3rd International Workshop on Research in the Large. 15-18.

[5] Higgs, M., Morrison, A., Girolami, M., Chalmers, M. (2013). Analysing User Behaviour Through Dynamic Population Models. To appear: CHI WIP, 2013.

[6] Kjeldskov, J., Skov, M., Als B., Høegh R. (2004) Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. Mobile Human-Computer Interaction - MobileHCI 2004, 61-73.

[7] Brown, B., Reeves, S., Sherwood, S. (2011). Into the wild: Challenges and opportunities for field trial methods. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI)*: 1657-1666.

[8] Rooksby, J. (2013). Does Professional Work Need to be Studied in a Natural Setting? A Secondary Analysis of a Laboratory Study of Software Developers. In Petre M and van der Hoek (eds.) *Software Designers in Action. A Human Centric Look at Design Work*. Chapman Hall/CRC Press. *Forthcoming*.

[9] Schön, D. (1983) *The Reflective Practitioner, How professionals think in action*, Basic Books.

[10] Sharrock, W. and B. Anderson (1994): The User as a Scenic Feature of Design Space. Design Studies, vol. 15, no. 1, pp. 5-18.

[11] Morrison, A., McMillan, D., Reeves, S., Sherwood, S., Chalmers, M. (2012). A Hybrid Mass Participation Approach to Mobile Software Trials. Proc. CHI '12, 1311-1320.

[12] Randell, R., S. Wilson, Woodward, P. (2011). Variations and Commonalities in Processes of Collaboration: The Need for Multi-Site Workplace Studies. Computer Supported Cooperative Work (CSCW), vol. 20, nos. 1–2: 37–59.

[13] Suchman, L. (1997). Centers for coordination: A case and some themes. In Resnick, L. B., et al. (eds.): Discourse, tools, and reasoning: Essays on situated cognition. Berlin: Springer-Verlag,

[14] Karasti, H., K. S. Baker, and F. Millerand (2010). Infrastructure Time: Long-Term Matters in Collaborative Development. Computer Supported Cooperative Work—An International Journal, vol. 19, nos. 3–4, pp. 377–415.

[15] Star, S. L. and K. Ruhleder (1994). Steps Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. In Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW'94), Chapel Hill, NC, USA, October 22–26 1994. New York: ACM Press, pp. 253–264.

[16] Pollock, N. and R. Williams (2010). e-Infrastructures: How Do We Know and Understand Them? Strategic Ethnography and the Biography of Artefacts. Computer Supported Cooperative Work (CSCW), vol. 19, no. 6, pp. 521–556.

[17] Martin, D. Rouncefield, M**.** Sommerville, I. Rodden, T. Viller, S. (2001). Finding Patterns in the Fieldwork**.** ECSCW'01 Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work. 2001: 39-58.

[18] Blomberg, J., Karasti, H. (2013). Reflections on 25 Years of Ethnography in CSCW. Journal of Computer Supported Cooperative Work. In Press.