

The Role of a Cohort in the Design and Evaluation of Pervasive Systems

Tim Jay, Danaë Stanton Fraser
Department of Psychology
University of Bath
Bath, UK
BA2 7AY
{t.jay, pssds}@bath.ac.za

ABSTRACT

In this paper, we describe a new methodology for the design and evaluation of pervasive systems. We have recruited a cohort of 30 participants who are engaging with an interdisciplinary pervasive computing project Cityware over 3 years. The cohort has been selected in order to represent a broad mix of ages and technological abilities so as to increase the ecological validity of evaluation of systems and applications developed in Cityware. We discuss some of the techniques and methods that we have been able to employ as a result of maintaining this group of participants and illustrate how their data feeds into Cityware studies and applications. While the costs of recruiting and maintaining a cohort such as this are relatively high, the benefits in terms of the depth, richness and validity of results produced in this way are considered to be significant. We discuss this in terms of the potential for new designs, the type of data one can collect and engagement within the city.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/Methodology*

General Terms

Design, Human Factors

Keywords

Design, Pervasive Systems, Methods, Techniques, Evaluation, Cohort, Mobile

1. INTRODUCTION

There are a number of serious shortcomings in current HCI evaluation. There is often a bias towards the use of laboratory testing of systems and applications [9]. Laboratory evaluation typically involves technologically literate participants in controlled conditions, limiting the realism and generalisability of findings. These problems are emphasized where there is no clear boundary between the system being evaluated and the context in which it is situated. Where methods outside the laboratory are employed, participants still often have high levels of technological ability, diminishing ecological validity. Mobile and pervasive systems are both, by their nature, difficult to

separate from the contexts in which they operate. These systems are also usually being designed for eventual use by a general public who are not necessarily technologically able.

Recent projects which have explored the relationship between pervasive systems and urban environments (e.g. the Equator IRC, Mobile Bristol, Urban Tapestries, Intel's Urban Atmospheres) have conducted their research largely through 'experiences' or 'performances' that in many cases cover relatively small physical areas and in almost all cases are held over short timescales. While previous studies have explored people's social behaviour and relationships with urban space and pervasive technologies and have provided useful findings on which we can build [1,3,4,6], they have not addressed the methodological challenges of longitudinal, city-scale studies.

An alternative approach, an example of which will be described in this paper, is to recruit a cohort of participants from the general public, with varying degrees of technological familiarity and ability, for the duration of a large-scale project. The cohort is involved in the design and evaluation of systems and applications using both qualitative and quantitative methods. While it may seem that the recruitment and maintenance of such a cohort could be a difficult or expensive operation, we aim to show that, especially in the case of a long-term, multi-faceted project, that the benefits in terms of validity and reliability of results far outweigh any costs.

Our work can be situated within a cooperative design framework which originated in Scandinavia involving adults in co-design techniques [2,7]. This approach was adapted for technology design with children and fuelled debate around engaging children as full design partnerships [5] or just engaging them in parts of the process [13]. Our approach is taking this design methodology with both adults and children and extending it for the design of pervasive technologies with a group of residents of the city. We would argue that our approach takes this form of design one step further, as rather than just collecting a set of requirements we are taking our cohort on a 'technology journey,' gaining their insights but additionally encouraging them to come up with future designs.

Our cohort is providing baseline data engaging in usability testing and trials. We are carrying out a rolling analysis which regularly integrates data collected through interaction with the cohort. Our method is evolving according to the contingencies of long-term field research, but we are drawing on and building on existing participatory methods that we have developed in previous work [14]. This longitudinal work is also complemented by the space syntax analysis carried out at the Bartlett and statistical data from Vodafone on user volume and movement in the city we are studying, Bath, a heritage city in the south-west of the UK.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DIS 2008, February 25-27, 2008, Cape Town, South Africa.
Copyright 2008 ACM 978-1-60558-002-9/08/0002...\$5.00.

This paper provides a short description of the Cityware project, followed by a rationale for the recruitment of the cohort and the work we have been engaging in with them. We present some of the baseline data we have collected from the cohort and then using two example applications, the context logger and Bluefish, discuss the novel types of data we have been able to collect.

2. THE CITYWARE COHORT

The Cityware project is a research collaboration between the University of Bath, Imperial College, The Bartlett at University College London, Vodafone, Nokia and Hewlett-Packard (<http://www.cityware.org.uk>). The main goal of the project is to increase our knowledge and understanding of people's relationships with urban spaces and with pervasive technologies in order to enable the development of tools and techniques for the implementation of long-term, city-scale pervasive systems. Evaluation is a core component of Cityware. The scope of the project is such that longitudinal evaluation is desired in addition to discrete, one-off studies. The longitudinal work focuses on how people perceive the city, how people use pervasive systems and how perceptions of the city and uses of pervasive systems interact with one another.

Within the project we are also carrying out targeted studies, deploying applications over relatively short periods (days to weeks) and analysing their usability and impact. However, our focus here is on our methodology for longitudinal design and evaluation which involved recruiting a cohort of participants for study and engagement throughout the duration of the project. We aim to have thirty participants engaged with the project over three years. A reserve cohort of 30 people has also been established, as has a mailing list of other interested members of the public.

The key novel aspect of this research is the sustained and longitudinal nature of our design and evaluation. The initial recruitment of the cohort was therefore crucial to the project.

2.1 Recruitment

It was important to consider, before advertising for participants, how the cohort should be made up. The first of our considerations had to do with the size of the cohort. The cohort consists of thirty participants – as many as possible within certain constraints, having to do with the manageability of data produced and the availability and cost of necessary equipment. A representative sample was deemed unnecessary, but it was desirable that the cohort should offer opportunities to explore group and individual differences in behaviour. The main factors we considered in determining the makeup of the cohort, given its size, in order to maximize these opportunities, were age and gender, occupation and home/work location.

Age diversity within the cohort is important as we expect younger people to interact with both the city and with technology in different ways to older adults. Usage of a mobile device is likely to be substantially different between younger and older adults. Nicholas and Chivhanga [11], for example, found that 43% of adults aged 45 and over never send an SMS message, compared with just 4% of those aged between 16 and 24.

Gender differences are also well represented in the literature [e.g. 8, 10, 15]. For males, the mobile phone offers an opportunity to invite the 'public world' into their 'private world', while for females, the mobile phone is an extension of their public and family life [10]. An alternative perspective is

offered by Katz [8], whereby for females, mobile phone use is associated with personal security and the maintenance of relationships with family and friends, while for males, phone use is associated with fashion, power and virility.

We endeavoured to include people with a range of occupations. Other population variables, such as ethnicity and economic status were considered but were not manipulated in the make up of the cohort. Additionally the proximity of participants' homes to the city centre was also considered, the aim being that the systems developed should be usable by both regular and infrequent visitors to Bath, and we therefore wanted the cohort to represent a range of familiarity with the city. (While we do not report on this within this paper it is worth noting that we are also carrying out studies with tourists to Bath).

In order to recruit the cohort in February 2006 a local press release was designed asking for participants to take part in the project. There was a high response rate with 200 respondents within a week. A high proportion of these respondents were male (80%) and the majority fell between 25-55 years of age. They responded from a variety of sources including The Bath Chronicle, The Register, IT Weekly and the University home Page. In order to gain a more diverse age range the Bath Chronicle ran a follow up piece encouraging the older and younger participants to apply. Unfortunately 'Cyber Guinea Pigs Still Needed' was not our choice of title! However it was successful. In order to sift through the 200 plus applications we sent out a questionnaire on use of technology in order to gain a feel for enthusiasm, with a strong caveat that we were not selecting based on technology expertise. We then shortlisted on our criteria age, gender, occupation and home/work location. Our final 30 cohort members included 10 females and 20 males. Occupations were varied including a foster carer, a magician, a fireman, golf course manager, pensions analyst, a pilot and a potter. Three of the cohort are under 18 years of old and needed parental and school consent. We also created a reserve cohort consisting of 30 people who receive regular newsletters about the project and all the other respondents are kept on file as potential participants.

All of the cohort were visited individually in order to discuss the project and were required to sign an IP disclaimer. The cohort is now an integral part of our team and have taken part in six workshops, an experimental study and also undertaken a number of challenges.

The primary purpose of the cohort is to participate in the iterative design and evaluation of applications and systems developed in Cityware. As the cohort will therefore be familiar with applications and systems developed in Cityware the role of the reserve cohort will be to take part in evaluation studies towards the end of the project where it may be important that participants are new to the application or system under investigation. Members of the reserve cohort can also be called upon to replace drop-outs from the main cohort.

There are some incentives for members of the cohort. We aim to make events involving the cohort as interesting and engaging as possible. We are also providing each member of the cohort with a mobile phone (currently a Nokia N70) and a Vodafone SIM that allows free unlimited voice, SMS and data transmission.

Part of the recruitment process involved the collection of information regarding participants' lifestyles and technological abilities. This has helped to ensure that we are working with a group that covers as broad a demographic as possible.

2.2 Data Collected

Initially, we gathered information on the cohort's lifestyle, perceptions of the city, technology use and trust in technology, using workshops, questionnaires and interviews.

There are three main areas of people's behaviour that we are interested in. These relate to relationships with the city, relationships with technology and associations between technology and the city. We are interested in how people relate to their neighbourhood and to the city. For example, we are studying how participants perceive and use landmarks, and how these are represented in various formats. Our understanding of people's relationship with space will feed into our construction of design guidelines for city-wide pervasive systems. From the reverse direction, we are interested in how participants currently use technology, particularly the mobile phone that they have been provided with as part of the project. The conjunction of technology-use and use of neighbourhood and city will provide an ideal dataset for informing system design.

A variety of methods have been employed during the project in order to gather data. Amongst these have been questionnaires, interviews and focus group work. We have also been using map-sketching and moblogging in order to learn about participants' perceptions of their neighbourhoods. A significant advantage of maintaining a cohort of participants is the ability to revisit questions at a later date in order to explore issues further on the basis of initial analysis. This makes it easier to triangulate data in order to address a problem from multiple perspectives.

We have a large amount of background data for the cohort. We know what phones participants were using before the start of the project. Participants have completed questionnaires regarding phone use at the beginning of the project, including functions and applications most commonly used.

The baseline data we have collected also includes personality information and data on use of technology. All participants have completed the Eysenck Personality Inventory. This gives scores on two scales, extraversion and neuroticism. We are exploring ways in which personality factors such as these relate to behaviour in relation to technology. We expect to see an association, for example, between high levels of neuroticism and concern for security and privacy of personal data. In addition, therefore, we have devised questionnaires to explore general technology use and attitudes towards privacy and security in respect of technology use.

A more specific example of data we have collected relates to participants' use of Bluetooth. An initial questionnaire on the subject of Bluetooth, answered by 55 respondents (including 30 cohort members), revealed that 94.5% of respondents know what Bluetooth is and 5.5% don't know.

52.7% only switch Bluetooth (BT) on when they want to use it compared to 16.4% who always have it switched on and 3.6% who simply don't use it.

21.8% of respondents leave their BT discoverable compared to 34.5% who only have discoverable BT when they want to use it and 18.2% who never have BT discoverable only paired. 10.9% have Bluetooth always switched on and always discoverable.

This kind of data is useful for informing our work in the development of context- (particularly location-) aware services. We are focussing on Bluetooth as a means for both detecting devices and for delivering content. Understanding Bluetooth

behaviour is a very important step towards the design of an effective system.

In a later workshop, we investigated Bluetooth use within the cohort further. Amongst those who regularly used Bluetooth (26 out of 30), 67% used it for file exchange only, 17% for voice calling (Bluetooth headset) and file exchange, 8% for voice calling only and 8% for file exchange and computer synchronisation.

We were also interested in naming behaviour. All Bluetooth devices are named, so that they can be easily identified when creating a connection. On most devices (including on mobile phones) these names can be changed. Within the cohort, 75% of participants had changed the name from the factory preset ('Nokia N70'). Some interesting findings regarding naming behaviour resulted from this workshop:

- 40% always use the same name (for all mobiles, computer, PDAs etc).
- 26% always use their own name
- 17% choose a name depending on current mood/how they are feeling at the time
- 9% choose a name depending on the company they are currently keeping
- 9% choose a name in the hope that someone will contact them

Names given to devices covered a wide range. Participants often used their own name or a nickname. Some examples of names chosen for mobile phones amongst the cohort, with reasons given:

- Moshi Moshi - Japanese for hello and because it's cool
- Magic Man - profession; owner is a magician
- Osama Bin Laden - to see if anyone noticed!
- Mudslinger - owner is a potter

All of the baseline data that we have for the cohort can be compared at a later date. This gives us a great deal of context for any data that we collect during the project, allowing for a very fine-grained analysis of behaviour. At the same time as participating in discrete design and evaluation studies, the cohort will be generating long-term datasets describing their use and experience of the Nokia N70 and the city of Bath.

3. CONTEXT LOGGER

3.1 The Mobile Context Logger Application

The context logger provides an excellent example of the benefits of a cohort. As has been described above to some extent, we will be asking for large amounts of data from the cohort, over a relatively long period of time. In order to achieve this, given the wide range of technological literacy within the group, we are employing a novel means of data collection. A context-logging application has been developed (by M. Huebscher and N. Dulay at Imperial College, London) and has been installed on the phones of the cohort. This automatically logs various data on the phone and forwards it, using the 3G/GPRS network to a server. The data collected include:

- Time, date and sender/receiver of text messages
- Time, date and sender/receiver of voice calls

- Camera images
- Applications in operation
- Location (in conjunction with a Bluetooth GPS unit)
- Bluetooth scans
- Cell ID in use

The context logger has been developed through testing with the cohort. This process has involved a gradual expansion of the context logger user group through an iterative process of improvement.

Initially, the application was tested for robustness by members of the research team. This allowed us to begin testing the application with members of the cohort without risking the disruption of participants' phone use as a result of operating system failure or other errors.

Following this, the context logger was introduced to four of the cohort. There were two areas that we were most interested in developing our understanding. The first area we were interested in was understanding how participants would make use of the context logger. The second was in understanding participants' reactions to the application in terms of privacy.

Two of the participants in the cohort declined to have the context logger installed on their phones, due to privacy concerns. This gave us an ideal opportunity to investigate the nature of these concerns. The two participants shared a major concern, namely that they had no control over what data was being sent, or when. In general, they were happy to have us observe data generated during their use of the phone. However, both felt that there were times when they would not want some data to be transmitted. The data that they were most concerned about were location and photos. One was also concerned about Bluetooth scans, as these would indicate co-presence with other identifiable people. This participant was also concerned that the content of voice and/or text messages could be logged, also that incoming data (e.g. photos/messages sent by friends and family) could be logged and transmitted.

This feedback was extremely useful to have at an early stage in the development of the context logger, before considering its deployment to a wider audience. In response to the comments we received we were able to make adaptations to the design of the application so that introduction to larger numbers of the group was possible. We were able to assure both participants that some of their concerns had already been accounted for in the design of the application – voice and text message content were not to be logged, for ethical reasons, nor were any incoming data (any data not originally generated on a participant's phone). In response to their need for control over the data logged, a user interface was designed that could be used in order to determine which data would and would not be logged and/or sent to our server at any given time.

While the negative feedback we received was useful for the development of the context logger application, the positive feedback that we received shed some light on its potential both as a personal data recorder and as a research tool. One participant, for example, regularly (at least once a week) reviews his logged data. As could be expected, he is particularly interested in reviewing logs of his location data, which can be viewed as trails on Google Earth. The context logger website allows the creation of a Google Earth mashup, including photos taken at locations on a journey. This has helped him to create a record of a holiday, for example (see Figure 1).

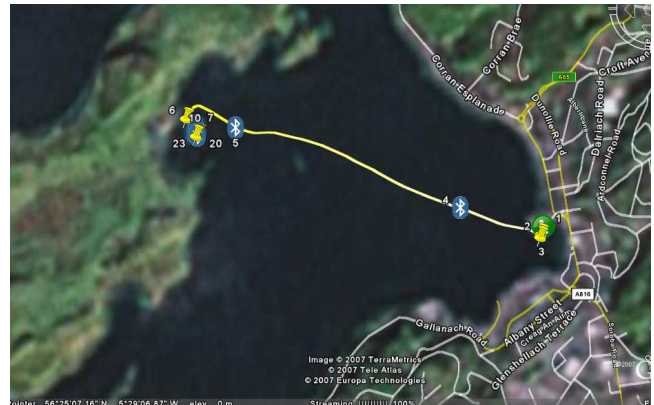


Figure 1: GPS trail produced during a sailing holiday in Scotland

Perhaps more unexpectedly, he regularly reviews his Bluetooth scan data. This is done purely out of interest, to see what Bluetooth names have been copresent with him during the day. He is interested both in the number of Bluetooth devices present in various locations that he visits and in the names that people choose to give their devices.

The fact that the context logs are interesting enough for participants to review on a weekly basis is promising for our research. As already discussed, it is very important for the continued success of the project that participants feel involved and engaged with our research.

Following adaptation made to the application in order to alleviate privacy concerns, the context logger is now installed on 20 of the participants' phones. We are now able to work on developing the context logger as both a research tool and a means of personal data logging.

3.2 Using the Context Logger

Any data that is held on the phone is potentially accessible to the logger. Therefore we can combine the location data and the phone use data to generate a picture of participants' use of technology related to their lifestyle and activity. The data generated by the cohort will feed into various Cityware activities. These include:

- Generating a picture of digital activity in the City of Bath
- Developing context-aware systems and applications for mobile phone users
- Investigating individual differences in use of technology
- Investigating relationships between technology-use and perceptions of the City of Bath and participants' neighbourhood

The context logger serves multiple purposes in Cityware. It will be used for transferring data for either short- or long-term targeted studies, as in our work regarding perceptions of neighbourhood and of the city of Bath. It can be used in our longitudinal study of interactions between perceptions of neighbourhood and city and technology use. It will also be used in the development of context-aware systems, whereby machine learning techniques are employed in order to generate and test predictions of behaviour based on context-logs, prior to the development of suitable applications. The context-logger will also be used in order for the cohort to feed into our database of Bluetooth activity in the city of Bath. When the context-logger is recording both local Bluetooth activity and location data,

they are effectively acting as mobile Bluetooth scanners – adding to the data described above, generated from static scanners placed around the city of Bath.

3.3 Perceptions of Neighbourhood and City

Our work in understanding relationships between participants’ perceptions of neighbourhood and city and their technology use will help to highlight the benefit that a cohort of participants brings to Cityware. The novel aspect to this work is the fact that we are able to work with the cohort for a great deal of time and therefore coordinate a number of different sources of data.

We are exploring perceptions of space over the three years of the project, using questionnaires, map-sketching activities and neighbourhood tours – each of which is repeated through the project in order to build a picture of changing perceptions.

There are two types of map-sketching activities that participants take part in. On a blank sheet of A4 paper, they are asked to draw a rough sketch map of their neighbourhood (or of the city of Bath). We deliberately do not give an explanation of what we mean by the term ‘neighbourhood’, we want to find out what participants themselves think it means for them. Participants are also asked to label anything that they consider to be an important feature on their map.

The second activity involves drawing a boundary on a pre-printed map of an area (again, either their neighbourhood, or the city centre). They are asked to draw a line representing the edge of their neighbourhood (or the edge of the city centre). They are also asked to mark and label landmarks within the boundary that they have drawn.

These maps communicate participants’ perceptions of space very effectively, along with the parts of those spaces that they feel are important or personal to them. However, the maps are not a perfect source of data. For example, it is difficult to know whether the maps drawn by participants are an accurate reflection of their use of the spaces represented. It is also difficult to know why maps are drawn as they are, or why a boundary follows a certain path, for example.

We are able to learn a lot more about people’s perceptions of space by asking them to take a tour, as if showing around a newcomer, using their mobile phone to record the experience. The mobile phone can be used to take still images, video, audio and text notes. The route of the tour, and the location at which each document was created, are logged automatically by the context-logger and can be sent directly to a remote server, avoiding difficulties of data transfer as a result of participants’ unfamiliarity with technology. This activity takes approximately an hour and can provide a very detailed picture of a person’s relationship with a space.

The context logger can also provide extremely useful background data regarding participants’ use of space. As it runs unobtrusively in the background on the mobile phone, it can be left to run semi-permanently, logging location. Over time, we can generate a picture of participants’ use of space, both in their neighbourhood and in the city.

These four sources of data – map-sketches, questionnaires, documented tours and long-term context logs – produce a rich and powerful picture of participants’ relationship with space. There are a number of questions that we are interested in answering, all of which will be answered more effectively as a result of the coordination and triangulation of these multiple datasets. These questions include:

- How do participants perceive their neighbourhoods?
- What do participants consider to be landmarks?
- How are landmarks used?
- What individual differences are there in perceptions of space?
- How are depictions of space similar or different across formats (e.g. map-sketch v. tour)?
- How does perception of space evolve over time?
- How does use of technology (i.e. moblogging, context logger) affect perception of space?

The long-term participation of the cohort, and deployment of the context-logger, will be invaluable in helping us explore the answers to these questions.

3.4 Preliminary Results

Results so far indicate some interesting themes for further investigation. Regarding neighbourhoods, it seems there that participants often have multiple conceptions of neighbourhood. An example from one of the participants from the cohort will help to illustrate this. We will draw on data from three sources – the sketch-map, the boundary drawing and the neighbourhood tour.

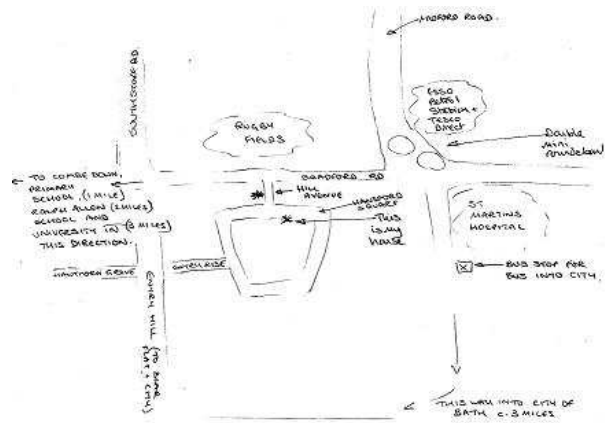


Figure 2: Map-sketch of Neighbourhood

This participant’s sketch-map covers a relatively small area. The participant’s house is in the centre of the map, and the edges of the map extend to cover an area approximately 3-4 minutes walk from the house (see Figure 2).

The boundary marking on the pre-printed map extends much further than 3-4 minutes walk, in order to include frequently used shops and services. The house is still at the centre of the neighbourhood as marked, but the shape of the boundary is irregular. For example, there is a loop to the north, in order to include a gym, and a loop to the south, in order to include a favourite walk to a local pub. See Figure 3.

Both the sketch-map and the boundary-marking task were introduced as having reference to the neighbourhood. Therefore, the difference in the area covered in each depiction of neighbourhood is interesting. The participant’s tour of her neighbourhood, recorded using her mobile phone (see Figure 4), sheds some light on the disparity between the sketch-map and the boundary marking activities. The route taken for the tour includes the loop to the south, taking in a favourite walk. In one audio note, she says that she considers the southern end of this loop to be part of her neighbourhood as it is a location

she often visits. However, on her return journey, back to her house, she makes another audio note, saying, “Now we’re coming back in to my neighbourhood”.

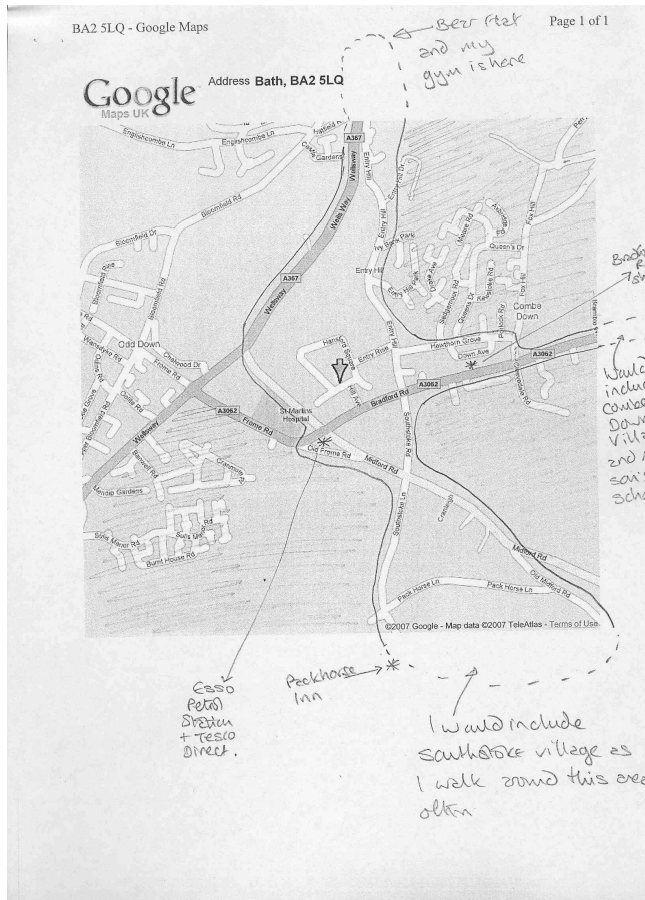


Figure 3: Sketching the Boundary of a Neighbourhood

The three sets of data help to demonstrate the two categories that form this participant’s perception of her neighbourhood. There is a neighbourhood defined by distance from home and a neighbourhood defined by use. The multiple sources of data available thus draw out aspects of perception that would otherwise be inaccessible.

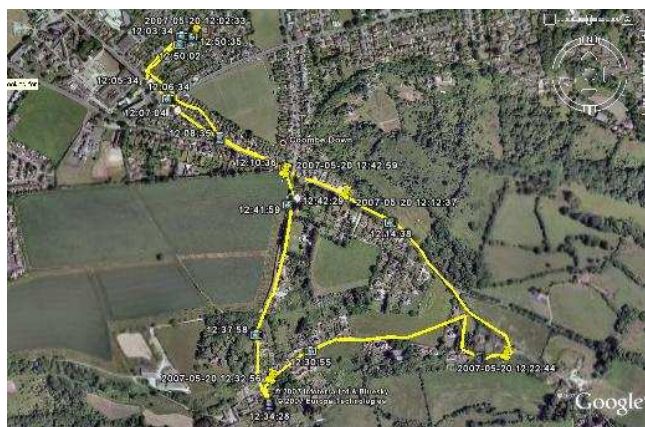


Figure 4: Google Earth Mashup of a Neighbourhood Tour

Results from other members of the cohort indicate a similar dual categorization of neighbourhood. In addition, we are seeing interesting patterns of landmark perception and use. Finally, we are noticing interesting differences in participants’ map-sketching before and after engaging in moblogging activities. Participants’ perceived size of their neighbourhood seems to grow after recording their neighbourhood tour.

The context logger, in addition to its use in recording one-off tours as shown above, provides us with long-term data regarding participants’ use of their neighbourhoods. We will be comparing this long-term data with the results generated through discrete tests to see how actual use of neighbourhoods compares to reported use.

We aim to follow up this work by expanding both the geographical area and the content that we are interested in. Firstly, we will be asking participants to take part in similar activities in Bath city centre. Secondly, we will be taking a greater interest in the ways in which the digital landscape (Bluetooth/Wifi activity etc.) plays a part in participants’ behaviour.

4. BLUEFISH

4.1 The Bluefish System

The ‘Bluefish’ system is a key component of Cityware. The aims of Bluefish are to make the digital landscape of the city of Bath visible to the public, in order to study public perceptions of pervasive systems and to explore the effect of these perceptions on behaviour.

The Bluefish system consists of a number of Bluetooth scanners and large display screens, placed around the city of Bath. The scanners are networked and feed in to a remote database in order to build a picture of digital activity in the city. The use within Cityware of Bluetooth scanners in describing digital activity in the city has been detailed elsewhere [12].

When a Bluetooth device is detected in the vicinity of one of the display screens, a representation of that device appears on the screen. This representation takes the form of a fish with the same name as the Bluetooth device, trailing a string of bubbles containing information concerning that device’s activity. This information consists of that device’s three most visited locations in the city along with the most often co-present device at each of those three locations.

4.2 Research Questions

The act of making this information public raises a number of questions about trust, privacy and security. In one sense, all of the information that will be displayed is publicly available. Any person with a suitable Bluetooth device, such as a mobile phone or notebook computer, can find out for themselves what other devices are in the vicinity. However, Bluetooth users do not expect this information to be made public in the form of a large-screen display, not do they expect their movements around the city to be tracked. The Bluefish system is conceived of as a privacy probe, designed to explore public perceptions of privacy and surveillance.

There are a number of questions that we are interested in addressing via the Bluefish system. Bluetooth activity is interesting because people can choose whether they want to broadcast their Bluetooth name or not, either by switching Bluetooth on or off, or by switching between ‘visible’ and ‘hidden’. We expect there to be competing influences affecting

Bluetooth behaviour in the vicinity of the screens. We expect the displays to be engaging for the public – causing some people to switch on their Bluetooth devices or cause them to be discoverable, in order to take part in the Bluefish system. On the other hand, we expect some people to have concerns surrounding privacy and surveillance – causing them to either switch off the Bluetooth facility on their device or otherwise hide its transmission. We want to understand the nature of both public engagement and public concern in respect of the Bluefish installation and to examine the effect of these perceptions on behaviour.

4.3 Evaluation Methods

The questions described above are in part answerable using the data generated by the system itself. All of the data generated by the scanners will be logged in a database. The scanners will be gathering data for a period of time before the installation of the display screen takes place. This means that we will be able to compare Bluetooth activity post-installation with pre-installation activity. Analysis of these datasets will allow us not only to monitor any increase or decrease in overall activity, but also to identify individual devices that have either ‘disappeared’ or ‘appeared’ in the city concurrently with the installation of the display screens. If, for example, a device has been consistently observable for several weeks up until installation and then, following installation, is no longer observable, there is a reasonable probability that the person to whom the device belongs has chosen to make it undiscoverable by our scanners. Conversely, if a device has not been observed by the system prior to installation, but is consistently observed for a period of time post-installation, there is a reasonable chance that the behaviour of the owner of that device has been influenced by the Bluefish installation.

Although it will be possible to learn a great deal through analysis of logged data, there are substantial aspects of our questions that cannot be accessed directly within the system. The behavioural analysis accessible via system logs can not tell us in any detail about people’s feelings of engagement, interest, concern, fear, annoyance or embarrassment. It is important that we understand why behaviour changes in response to the system. This is the kind of situation where a committed cohort of participants, engaged with the project and representing a wide range of ages and levels of technological knowledge becomes extremely valuable.

It would be possible to conduct some evaluation of the Bluefish system by recruiting members of the public in the vicinity of the screens. Questionnaires and/or short interviews would help shed some light on behaviour in response to the Bluefish system. However, this process would be very inefficient. We know from previous work [12] that the number of people carrying a discoverable Bluetooth device falls somewhere between 7% and 10% of the pedestrian population of the city of Bath. Therefore we could expect to have to approach ten people in order to find one person who regularly uses a Bluetooth device. Compounding this issue is the fact that we expect that people engaged with the system are likely to spend much more time near the display screens that people who are not interested, or have privacy concerns. It is those people who have concerns surrounding the system that are likely to be the most interesting to talk to, but the least accessible.

If interviewing members of the general public, it could also be difficult to gauge the validity of responses. This would be especially so in the case where a person has turned off their Bluetooth device as a result of concerns over privacy. If a

person makes their Bluetooth device undiscoverable as a result of the Bluefish installation, they are somewhat unlikely to reveal that information in an interview.

The benefits of access to the cohort are best appreciated when we consider the richness of data we have regarding their perceptions and behaviour in relation to the city and to technology over the duration of the project. When they tell us, through focus groups, interviews and questionnaires, of their responses to the Bluefish system, we can see how those responses relate to their responses in other situations. For example, we have data from a general technological security questionnaire, previously administered to the cohort, which contains questions about security in relation to the use of home computers and mobile devices and also about feelings regarding surveillance in connection with citywide CCTV networks and national ID card schemes. We also have data from personality scales administered to the cohort. These data add a great amount to the usefulness of responses to the Bluefish system.

The comparison and coordination of responses to Bluefish, accessible through system logs, interviews with the general public, and activities with the cohort will have a powerful effect on the validity of our evaluation. Each source of data will serve to reinforce and explain the other two.

5. ENGAGEMENT AND RETENTION

It is important for the project that as many participants as possible remain members of the cohort for as long as possible. There are a number of measures that we have taken and will be taking to keep rates of attrition low.

During the recruitment of the cohort, it was considered essential to find individuals who were committed to the project. It was important to make clear to the cohort what demands the project would make of them. These would include regular meetings and workshop sessions, and the disclosure of various data regarding their lifestyle and behaviour. We aim to have meetings in the evenings in order to involve everyone and build a team environment.

Over one year into the project, we have had four out of the original thirty members of the cohort drop out. The time commitment – attendance at meetings and workshops – has been the main reason for dropping out of the cohort.

The draw of the free phone and SIM card is still a good incentive for the cohort, but possibly not to the degree that it was a year ago. The Nokia N70 is no longer at the forefront of mobile technology, and we have had several inquiries from the cohort regarding possible upgrades. It seems likely that we will need to replace the phones with newer models in the near future in order to retain as many participants as possible.

We are finding that a number of the cohort are now also being provided with a phone and SIM by their employer. This has two effects on the project – firstly, this again reduces the value of the free phone and SIM that we are providing – secondly, participants are less likely to use the phone we are providing as their main phone. The second issue will become more of a problem as we demand more data from participants. In order to develop the kinds of context-aware services that we are interested in, we will need to collect large amounts of context data. This will be difficult to achieve unless participants are using the Cityware phone as their primary mobile device. Again, this problem can be solved by ensuring that the technology we are using is up to date.

Despite the technology issue, the vast majority of the cohort are very positively engaged with the project. They are keen to suggest ideas for applications or alternative perspectives on the use of systems or the interpretation of data. This engagement has been achieved at least in part as a result of efforts to give the cohort some ownership of the project and the data generated. We aim to make the events and activities that the cohort is involved in as interesting and enjoyable as possible. This has the effect of maintaining an active, engaged cohort, in turn raising the quality of data generated.

It is important that we retain as many of the cohort as possible for the duration of the project, in order that we retain the longitudinal aspects of our work. However, comparisons between long- and short-term members are likely to inform our understanding of developing perceptions of and relationships between technology and the city.

6. CONCLUSIONS

There are three key aspects of the research that we want to emphasise in this paper. The first is the novel design and methodology employed. The second are the new forms of data collected and the combination of multiple data sources involved. Thirdly, there is the engagement of the city and community in the research.

6.1 Novel Design and Methodology

HCI research is often difficult to generalize as a result of a lack of realistic context in evaluation. More realistic contexts can be achieved using field studies, but these have the disadvantage of difficult data collection and high time-costs. Generalisability of HCI evaluation can also be compromised by the use of biased samples of participants. Samples often consist of Computer Science undergraduates or similar, who are likely to be familiar with technology and therefore not necessarily representative of the general public. Again, high costs are often associated with the recruitment of a more varied group of participants.

This paper has described the recruitment of a cohort of thirty participants, with a wide range of age and familiarity with technology. This group has been recruited for the duration of the Cityware project – a large-scale project with a 3-year duration. This level of sustained engagement over time is a novel way to solve the problem of ecological validity in HCI research.

The main advantages of maintaining a cohort are the ability to attack a question on several fronts, using multiple data sources, and the ability to ensure validity in the evaluation of systems and applications. The Bluefish system is an example of a situation in which it is important that participants involved in evaluation represent a varied population. We are primarily interested in using Bluefish as a privacy probe – we anticipate that the system will provoke reactions that will help us to explore perceptions of security and privacy. It is likely that technologically literacy will be a factor affecting participants' responses.

Bluefish is also an example of a situation in which the recruitment of participants in-situ would be extremely difficult. It would be possible to locate and question those people who were engaged with the system, but it is the people who are not engaged with the system that are likely to be the most interesting. Disengagement with the system amongst Bluetooth users will be a result of a lack of interest or of privacy concerns. It will be very difficult to, in the first place identify, and in the second place recruit, participants for interview who are

disengaged with the system. Given that we are interested in understanding the nature of privacy concerns in response to Bluefish, it vital that we avoid this bias.

Having a cohort of participants involved in Cityware helps us to solve both of these issues. The cohort consists of a wide range of participants who can be directed to engage with systems such as Bluefish in ways that suit the needs of the project. They represent a captive audience of potential users, whether engaged or disengaged with the system we are studying.

6.2 Combinations with Data

Perhaps the most valuable aspect of the use of the cohort, illustrated in part by our use of the context logger, is the integration of datasets to answer large and complex questions. The context logger, alongside the variety of background data that we have collected for the cohort, allows us to situate observed behaviour in the context in which it takes place. The reliability and validity of our analysis, taking the context into account, can be greatly increased.

A good example of our use of multiple data sources in addressing a research question is our exploration of participants' use of both their neighbourhoods and the city. On the one hand, we have data from discrete activities in which the cohort has taken part. The location data from participants' neighbourhood tours, along with documents such as camera images, video, text and audio notes together provide a surprising level of insight into people's relationship with space. As the logging and transmission process occurs in the background, participants are able to ignore the process and focus on the activity which they are engaged in, without having to focus on, or try to solve, technical issues. Map-sketches, boundary sketches and questionnaires also tell us about participants' perceptions of, and relationship with, their neighbourhood. On the other hand, we have long-term data, from the context logger, regarding participants' use of their neighbourhoods. We can compare the data we have from discrete activities with 24/7 location data over several weeks. This allows us to determine how participants' reported relationship with their neighbourhood is related to their actual use of their neighbourhood.

The use of multiple data source to address a question in this way will allow for a greater depth of understanding than could be achieved using traditional methods.

6.3 Engagement of a Community

The involvement of the city and the community in Cityware is extremely important. The aim of the project is to develop city-scale pervasive systems for the city of Bath. We want to provide something that the city of Bath can engage with. Without the input of the community that live and work here, this development would run a great risk of missing its target.

On one level, we achieve engagement with the city through our partnership with Bath and North-East Somerset Council. At a more fine-grained level, we achieve this through our work with the cohort.

Throughout the project we have made efforts to ensure that participants have some ownership of the project. They are a major part of both design and evaluation processes and feel valued members of the project. As a result, we are able to work with (mostly) the same group of participants throughout the lifetime of the project, and we have confidence that our development of systems is appropriate to the needs of the city.

6.4 Summary

There are some high costs involved in working with a cohort of participants for the duration of a project like Cityware, but there are also some more than worthwhile benefits. The main costs are the time and effort required to recruit and retain the right kinds of participants for the project. There has also been the need to develop novel methods of data transmission between cohort and researchers. The benefits, from our point of view, outweigh these costs. These consist in the ability to generate large quantities of focused data regarding the behaviours of the cohort in relation to both the city and in their use of technology. The cohort has a role in both the evolution (through iterative design processes) and evaluation of systems that we develop.

The main advantage of maintaining the cohort has been that we are able to reduce problems of generalisability associated with laboratory studies. Realism is maximized by selecting a group of participants in order to ensure a range of ages and technological abilities. The cohort is not intended to be a genuinely representative sample of the population, but to give us opportunities to explore individual differences in behaviours that we are interested in.

7. ACKNOWLEDGMENTS

The Cityware project is funded by the EPSRC WINES programme. We would like to thank our local team, Eamonn O'Neill, Tim Kindberg, Abha Joshi and Vassilis Kostakos and acknowledge their role, particularly in the development of Bluefish. We would also like to acknowledge the rest of the Cityware team, particularly Markus Huebscher and Naranker Dulay for their role in the development of the context logger. Thanks are of course due to the Cityware cohort, who have given a great deal of their time and thought to the project. Also to Tim Jones, who had a key role in their recruitment.

8. REFERENCES

- [1] Benford, S., Seagar, W., Flintham, M., Anastasi, R., Rowland, D., Humble, J., Stanton, D., Bowers, J., Tandavanitj, N., Adams, M., Row Farr, J., Oldroyd, A. and Sutton, J. The Error of our Ways: The experience of Self-Reported Position in a Location-Based Game. The 6th International Conference on Ubiquitous Computing. (UbiComp). Nottingham, September 2004. Springer/ACM pp. 70-87.
- [2] Bodker, S., Ehn, P., Kammersgaard, J., Kyng, M., Sundblad, Y. A Utopian Experience. Proceedings of the 1986 Conference on Computers and Democracy, Avebury pp. 251-278.
- [3] Brown B., Chalmers, M., Bell, M., MacColl, I., Hall, M., MacColl, I., Rudman, P. Sharing the square: collaborative leisure in the city streets. Proceedings of ECSCW 2005, Paris, France. Springer. p427-429
- [4] Cheverst, K. and Schmidt-Belz, B. "Mobile guides – an HCI perspective", editorial, Special Issue of Journal of Behaviour and IT on 'HCI in Mobile Guides', Guest editors, K. Cheverst, B. Schmidt-Belz, Taylor and Francis Ltd. Volume 24, Number. 1, Jan-Feb 2005.
- [5] Druin, A., Cooperative Inquiry: Developing new technologies for children with children. Proceedings of the ACM SIGCHI 99 Conference on Human Factors in Computing Systems, Pittsburgh, USA, 15-20 May, pp. 592-599
- [6] Fraser, M., Stanton, D., Ng, K. H., Benford, S., O'Malley, C., Bowers, J., Taxén, G., Ferris, K., Hindmarsh, J. Assembling History: Achieving Coherent Experiences with Diverse Technologies. In Proceedings of ECSCW 2003, Helsinki, Finland. Kluwer. pp.179-198
- [7] Greenbaum, J. and Kyng, M., Design at work: Cooperative design of computer systems. Hillsdale, NJ: Lawrence Erlbaum. 1991
- [8] Katz, J.E. Connections: Social and cultural studies of the telephone in American life. New Jersey: Transaction Publishers, 1999
- [9] Kjeldskov, J. and Graham, C. "A review of mobile HCI research methods," in Human-Computer Interaction with Mobile Devices and Services: 5th International Symposium, Mobile HCI 2003 Udine, Italy, Lecture Notes in Computer Science. pp. 317-335.
- [10] Lemish, D. "On the gendered nature of mobile phone culture in Israel," Sex Roles, pp. 511-521, 2005.
- [11] Nicholas, D. and Chivhanga, M. "The communications love-affair: Children and the mobile phone," Library Association Record, vol. 104, pp. 90-95, 2002.
- [12] O'Neill, E., Kostakos, V., Kindberg, T., Fatah gen. Shieck, A., Penn, A., Stanton Fraser, D., Jones, T. "Instrumenting the city: Developing methods for observing and understanding the digital cityscape," Proceedings of the 8th International Conference on Ubiquitous Computing, 2006.
- [13] Scaife, M. and Rogers, Y, Kids as informants: Telling us what we didn't know or confirming what we knew already, in A. Druin (ed.) The Design of Children's Technology. San Francisco, CA: Morgan Kaufmann. 1999
- [14] Stanton, D., Bayon, V., Neale, H., Ghali, A., Benford, S., Cobb, S., Ingram, R., Wilson, J., Pridmore, T., O'Malley, C. Classroom Collaboration in the Design of Tangible Interfaces for Storytelling. In Proceedings of Human Factors in Computing Systems (CHI 2001), ACM Press, Seattle, Washington, pp. 482-489
- [15] J. Vincent, '11-16 mobile': Examining mobile phone and ICT use amongst children aged 11-16: University of Surrey: Digital World Research Centre, 2004.