

# Evaluation Probes

Anna Luusua<sup>1</sup>, Johanna Ylipulli<sup>2</sup>, Marko Jurmu<sup>2</sup>, Henrika Pihlajaniemi<sup>1</sup>, Piia Markkanen<sup>1</sup>, Timo Ojala<sup>2</sup>

<sup>1</sup>Oulu School of Architecture, University of Oulu, Finland

<sup>2</sup>Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland  
 firstname.lastname@oulu.fi

## ABSTRACT

We introduce *evaluation probes* for conducting emic, experiential evaluation of urban technologies “in the wild” without direct researcher presence. We commence with a thorough discussion and analysis of the original cultural probes, used by Gaver, Dunne and Pacenti to gain design inspiration, and their subsequent variations. We develop the concept of evaluation probes through careful re-conceptualization and application of the cultural probes in three successive studies conducted in the wild. We recount and reflect on our use of evaluation probes and discuss their merits and limitations in experiential emic evaluation.

## Author Keywords

Methodology; evaluation; in the wild; probes; experience.

## ACM Classification Keywords

J.5 Arts and humanities; J.4 Social and behavioral sciences.

## INTRODUCTION

The experiential evaluation of interactive installations “in the wild” [41] remains a challenge for the urban computing [28] research and design community. As computing becomes ever more pervasive in urban environments, researchers and designers of urban computing applications have become aware of the need to evaluate their works in the wild, i.e. outside of traditional research laboratories [7]. This work has already begun and much research has been conducted to evaluate interactive urban computing systems [27, 32]. However, those approaching the task of evaluative research from an *experiential* point of view [27] have stated that there is a lack of methods in the field. This is no task for strictly itemized check-lists; rather, there is a need to be able to flexibly gauge people’s holistic experiences of what it is like to come face to face with interactivity in public urban places. We agree with Rogers [41] that a “central part of designing in the wild is evaluating prototypes in situ. This involves observing and recording what people do and how this changes over suitable periods of time”. Researchers’ *etic* perspectives on people’s behaviour are important; but we also argue in favour of studying people’s *emic* experiences with interactive technologies. *Etic* refers

to researcher’s own interpretations and categorizations of people’s behaviour, whereas *emic* refers to insider knowledge: how people explain and categorize their own experiences. [3]

The task of developing these methods is not easy, as urban computing as a research field could be seen as involving several existing disciplines, and has strictly post-positivist roots [1, 21, 22]. Yet research questions may come in any form, and they require appropriate research methods in order to be answered. Issues relating to human experience, we argue, seem to be an area where this problematic is particularly visible. There is a crisis of experiential evaluation [2] that it is leading, we argue, to a much larger challenge in the urban computing community. The piecemeal addition and borrowing of individual methods and theories has proven difficult, if not impossible, as there is no common basis or understanding from where to proceed as a community. This, among other things, has led some researchers [e.g. 30, 31] to lament the lack of completely original methods and theories in HCI related fields. However, there is no need to reinvent the wheel many times over. In-the-wild urban computing studies are an inherently multidisciplinary field that must borrow from older research and design disciplines. This is not at all detrimental, and neither is urban computing alone in this regard. In the field of architecture and urban design (which always operates in real-world contexts), a similar variance of research questions and intentions prevails, and researchers in the field must necessarily operate within different research paradigms [20]. This is not seen as a problem, however, but researchers must always carefully state their position in order to avoid confusion and misinterpretation of their research methods and results; and, be keenly aware that all methods always come with epistemological roots and associated practices, accumulated through collective experience. Once care is taken to understand and respect the origins of theories, methodologies and paradigms, we can begin to build upon them, thinking about them in new ways, and finding uses for them outside of their original contexts.

In this spirit, we address the challenge of emic experiential in-the-wild evaluation of urban interactive technologies by introducing *evaluation probes* as a novel adaptation of the probes methodology [17, 18].

## PROBES AS METHODOLOGY

*Cultural probes* were originally introduced by Gaver *et al.* [17] as a set of carefully designed objects that were

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea  
 Copyright 2015 ACM 978-1-4503-3145-6/15/04...\$15.00  
<http://dx.doi.org/10.1145/2702123.2702466>

personally given to study participants. Importantly, cultural probes were based on the philosophical thinking of *artist-designers*; they could be described both as design research [e.g. 14], and as design through research. This very broad research methodology can be seen as falling under the general paradigm of emancipatory research, as opposed to (post)positivism and naturalism [20]. The profound aim was to inspire design by initiating a creative conversation between elderly people and designers; the probes were interventions meant to provoke new ways of thinking and to stimulate the imagination of both parties by distancing them from routine thinking. Instead of analysing the elderly participants' lives to supply tools to fulfill their needs, the researchers were more interested in opening up new design spaces through creating novel pleasures for them. Further, Gaver *et al.* openly stated that their probes were not intended to be analyzed at all. This is understandable in the research by design paradigm, wherein the act of designing is considered a tool for knowledge production. Overall, their approach was characterized by somewhat rebellious, subversive, critical and reflexive attitude.

#### Need for epistemological consistency

Cultural probes have been subsequently employed in various ways, raising enthusiasm for their applicability, but also concerns over misuse [4]. The term 'cultural probe' has been used to refer to many different approaches which bear some kind of similarity to the original probes, e.g. 'design probes' [33, 34], 'technology probes' [26] and 'urban probes' [37]. As the originators of the cultural probes approach did not state their epistemological and ontological groundings in much detail, there has been some confusion over their appropriate use. For example, Graham *et al.* [19] stated that probes have been used within different communities of practice, with different notions of vocabulary, practices and rigour; they pointed out that these communities have very different ideas concerning the interpretation of probe data. This also explained their own confusion over the designerly sensibility of cultural probes; they claimed that the discussion around probes cloaked 'design' in mystery, did not acknowledge accountability, and presented a view of design work that was not truly interdisciplinary. In a similar vein, Hemmings *et al.* [23] criticized the original approach, observing that Gaver *et al.*'s "inspiration" was largely a misnomer for acquiring ethnographic understandings, i.e. information about participants. We find these criticisms problematic, since they fail to understand and empathize with the essence of designerly thinking [9] and design research [14], which is based on reflective practice [42]. However, we do acknowledge that design knowledge and practices can appear opaque to readers who are not personally familiar with conducting design processes, as much of design knowledge is tacit [40] and involves lateral thinking [5]. For those interested, the aspect of design thinking and design inspiration in the context of probes has been more thoroughly explicated by Mattelmäki [33].

A fundamental premise in many papers [e.g. 19, 43] is that, as researchers, we must explain *how probes work*. This demand, when taken as a requirement for a more careful description of how knowledge is actually gleaned from acquired research materials, is understandable from a researcher's point of view and it must be taken seriously. However, we argue that rather than passively take the probes simply as they are, as an immutable recipe, we must also actively decide how we want them to work and boldly and reflectively experiment with them. Still, we agree with Boehner *et al.* [4] that methods that draw on cultural probes but change some of their essential aspects "cannot rest on the common acceptance of cultural probes for their validity." As such, "adaptations of the original cultural probes should be grounded in an awareness of which essential aspects of those probes are being adopted and which are not, and should justify those decisions." We agree that using methods without paying any attention to the larger methodology can be considered dangerous. This is comparable to the situation that is very often present in mixed methods research [10]: "Research involves a wide range of methodological tools, and we have to use many of these tools in a concrete research project. In other words, there is often a need to mix methods. However, we argue that this mix cannot be done without taking the ontological and epistemological dimensions into account. We call this perspective a 'critical methodological pluralism'." We deem it important, then, that when taking an approach and using it for other purposes, we must explicitly reflect on: 1) what aspects of the original methodology we are using and what we are leaving out, 2) what does this imply, and 3) what kind of epistemological commitments the methodology has, and does it match with our epistemological premises. In other words, we consider it crucial for the study to explicitly reflect on its *epistemological consistency*.

#### Essential features and epistemological commitments

To answer these questions, we must first elaborate on what exactly were the central features of probes. Others have also strived to tease out their essence. For example, Boehner *et al.* [4] identified four ways in which probes have been viewed by the HCI community: probes as packets, probes as data collection, probes as participatory, and probes as a sensibility. Graham *et al.* [21] identified five functions of probes common to most studies: to capture artifacts, yield (auto)biographical accounts, make the invisible (i.e. thoughts, attitudes, ideas, practices) visible, frame the participant as expert, and engage them in dialogue and conversation with designers and researchers. We agree that these features are important for the success of probes.

However, we argue that at the very core of the popularity of the methodology is (1) *the powerful metaphor of the "probe"* that effectively conveys the basic idea of how this type of research is carried out: a probe is an *artefact* that is sent into inaccessible places and situations in the real world to advance a researcher's understanding of that place, usually by collecting data. What is significant about the

adoption of this metaphor is that it is used for qualitative explorations. A cultural probe bears some similarities to the *diary* method in social sciences [15] through which temporally organized written reflection is collected. While both can be described as participant self-documentation methods, probes are a more expansive and accommodating methodology, owing in large part to the difference in the two metaphors, i.e. diary vs. probe. Probes can contain multiple non-chronological tasks and they allow for various types of materials and modalities to be used, i.e. the use of various media, such as cut-and-paste pictures, photos, postcards and drawings. Further, we consider (2) *the aesthetically pleasing and playful presentation* of the original cultural probes as a central feature that made them work. Gaver *et al.* specifically mentioned that they wanted to make the probe package look like a gift, given to the participants as a gesture of appreciation in order to entice them to truly engage with it. This could take many forms, depending on the task at hand. (3) *Critical participant reflection* was also an inherent and important part of the original cultural probes, to help participants “think outside the box”, for example through the use of exciting and unusual exercises or materials. Supporting open-minded or even critical thinking is intended to yield better, more ethical and socially sustainable design ideas. This cannot be done without (4) *developing an adequate rapport with the participants*; as such, it can be considered a crucial part of the methodology. This relationship should ideally be one of (5) *mutual learning/knowledge-production*; i.e. the researcher directs the participants’ attention to something in their everyday life, and helps them reflect upon it through the probe materials. The materials in themselves function as an enabler and a reminder on the researchers’ behalf. Overall, probes require (6) *open reflection and a critical attitude* in both their design and their analysis (or non-analytic use, as in the case of Gaver *et al.*). It is, then, a typical qualitative method in the sense that the researchers’ skill and ethics play a key part in making sense of the results [20]. In our view, then, a probe is an artefact or a collection of artefacts that is intended to both help research participants generate experiential data through reflective assignments, and to collect this data as various kinds of empirical research materials from participants.

It would seem that most of the aspects attributed to probes in this article and in review papers fall under three major umbrella terms: either they describe the *physical probe* artefact, the *participation* that is inherent to them, or they refer to the *aims* of the probe (originally, of course, to gain design inspiration). These are guided by an overall critical and reflexive attitude that is typical to the epistemology of the emancipatory research paradigm [20]. Even though this is not the only way to try to grasp what is essential about probes, we find it nevertheless a very useful one, and for the purposes of our research, we use this framework (Fig. 1) to describe probes.

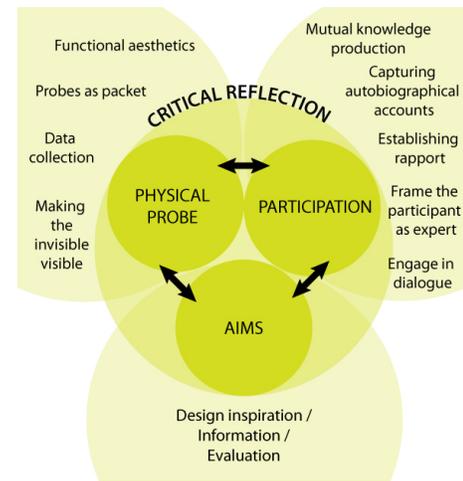


Figure 1. Conceptualizing the central elements of probes.

© Anna Luusua 2015

### EVOLUTION OF EVALUATION PROBES

Next we describe, in a chronological manner, how the new concept of evaluation probes evolved through our own work. An important first step in their development was the ICT diary probe, whose primary objective was to understand young people’s attitudes, perspectives and experiences about information and communication technology and its use. In other words, the overall aim of the probe was to gain information. We conceptualized the ICT diary probe as a hybrid, an evolutionary step between a cultural probe, a diary and an interview [15]. Realizing the vast potential of the approach for real-world evaluation, we developed it further into the concept of evaluation probes in two successive studies. These were “pure” evaluation probes in that enticing materials were linked with designed artefacts. The first was an evaluation of a network of outdoor public displays (Display probe). The second was an evaluation of a more short-term installation of adaptive lighting (Adaptive lighting probe). These experimental studies provide an interesting mix of research contexts and purposes within which probes were used. Their analyses have yielded valuable results detailed in [29, 44, 45, 46]. This paper focuses on the novel method that produced these results.

#### (1) ICT diary probe

The ICT diary probe study was carried out in Oulu, Finland; a Nordic city of approximately 141 000 inhabitants, in 2011-2012. In this study, we had a broad scope and mainly ethnographic aims; therefore we describe this probe as an information probe. The overall *aim* of the study was to map young adults’ everyday life with ICT in general, unveiling technology related attitudes, perceptions, experiences, skills and dreams. The motivation to employ a probe instead of more conventional ethnographic methods was the all-encompassing nature of ICT: it is used almost everywhere and in any situation, also in the home, which is a sensitive study site. A traditional ethnographic approach

combining participatory observation and interviews was deemed too intrusive and challenging.



Figure 2. Scrapbooks from the ICT diary probe study.  
© Anna Luusua 2015

Before the probe was designed, the responsible researcher conducted four group discussions with 20 young adults to gain preliminary understanding of the current ICT practices of the target age group. These discussions, carried out in an open and informal manner in local cafes, gave invaluable insights and affected the design of the probe. In addition, before recruiting the rest of the participants, the probe was tested with five persons and polished according to their comments. The final *physical probe* was a small, colorful scrapbook, accompanied with three-pages set of printed images and a return envelope (Fig. 2). It included ten different tasks, which required self-expression through cut-and-paste pictures, sketches and informal writing, tracing how and how much young adults use different ICTs; how they use them in different places, both at home and in public; what kind of attitudes they have towards the use and towards particular devices; and what kind technology-related hopes and fears they have. In addition, the diary included a task where participants were asked to go out and try one of the large interactive displays, located in different public places in the city, and write about their experiences. Importantly, this one simple and open-ended task became an inspiration for the later developed evaluation probes.

Concerning *participation*, participants were asked to keep the scrapbook at least for two weeks and return it to the researcher by mail. Then they were called to participate in group interviews that were conducted in a conversational spirit and structured around the tasks in the scrapbooks. The participants could reflect on the themes and compare their views which also made the differences within the age group very visible for the researcher. The interviews lasted for 1.5-2.5 hours and they were recorded and transcribed. In total, 48 participants filled in the scrapbook and participated in the interviews. The return rate was excellent as the probe was initially given to 56 participants; this was probably due to the fact that the researcher sent polite reminders by email every week and was ready to give extra time for the participants. Further, a majority of the participants reported that they had found completing the diary probe enjoyable. However, some also stated that it was not a conventional diary, reflecting its hybrid nature. The ones completing the

diary and interview were rewarded with two film tickets and a chance to participate in a lottery where they could win a gift card to a restaurant.

The returned scrapbooks were filled with essays, shorter notes, drawings, cartoons, sketches of floorplans and picture collages, and hundreds of pages of transcribed interviews. The “thickness” and breadth of the data have allowed us to look at young adults’ ICT reality from many different perspectives [29, 44, 45, 46]. This richness can be contrasted with the fact that collecting the material took a long time. However, it probably would not have been any faster to conduct several smaller studies.

## (2) Display probe

The Display probe was the first proper *evaluation probe* study, designed to evaluate participants’ experiences of a design artefact; in this case a network of public displays in an authentic urban setting. The research site was again Oulu, Finland, where a rich pervasive computing infrastructure has been deployed. This includes a network of 17 situated displays, which have been in operation since 2009. The displays are installed at pivotal indoor and outdoor locations, including the pedestrian area at the heart of the city. As such, these public displays are a familiar sight to the inhabitants of the city. Our study focused on the outdoor displays dispersed around the city center, i.e. at the pedestrian area and the nearby marketplace. At the time of their installation, the displays were a novel technology, and to this day, this network of public interactive displays remains the largest of its kind deployed for research purposes in a city center [36]. As such, it poses a wealth of unanswered research questions.

The *aim* of the study was to explore local people’s perceptions of the public displays, and how they experience the displays as a part of the streetscape in their everyday living environment from the point of view of aesthetics. The probe, then, can be described as an evaluation probe, as it was intended to specifically evaluate the experience of a particular design artefact. Further, the evaluated artifact, in this case a public display, is an integral part of the evaluation probe package; a necessary feature which sets it apart from cultural probes. This, in turn, affected other aspects of the probe; e.g. the design of the physical probe materials and the manner of participation. The probe was not meant to directly inspire design ideas, but rather to delve deeply into the experiences of our participants. As such, these probes were intended for in-depth qualitative analysis from their inception.

Participants needed to be able to carry the *physical probe* materials with them outdoors in winter and autumn conditions, even though they were not required to write on the spot. This is why we decided to gather all the material in a single notebook (Fig. 3). This, we have discovered, was a necessary practicality in the design of the evaluation probe, to enable participants to carry them comfortably. The content of the notebooks was carefully designed to

accommodate thoughts, opinions and ideas in an open-ended manner on several themes, e.g. the location of the chosen display, the display itself, and the general environment around it. Participants were instructed to express themselves freely through sketches, writings and photography. We specifically mentioned that the more creative types of writing, such as poetry, drawings and anything else they might deem useful, like cut-and-paste pictures, were allowed. One assignment specifically asked for visual input in the form of a photograph or a drawing. These proved useful, since a few participants clearly favoured visuals to text, and their sketches always conveyed thoughts that were different from their writings. Interestingly, the sketches and the one poem returned all conveyed *emotion*; specifically that the displays sometimes looked *lonely* out there on the streets. The probe methodology, then, not only seems to support different kinds of communication styles, but the different modes of communication seem to yield different kinds of knowledge. This makes the probes methodology more suitable to us than, e.g., the diary method, as we found the more expansive probe metaphor to be more conducive to participant creativity.



Figure 3. The Display probe notebook. © Anna Luusua 2015

In terms of *participation*, in accordance with our research agenda, we focused on young adults aged 20-29. The probe was conducted twice: the first probe in the winter (Feb-Mar 2013) had 20 participants and the second probe in the fall (Sep-Oct 2013) had 21 participants. The two probes conducted in different times of the year were motivated by our desire to evaluate seasonal effects on the participants' experiences. We handed out the notebooks individually, instructing and discussing the assignment with each participant. The participants seemed to be highly interested in our study and presented many questions about ourselves and our work. This, we argue, helped us to bridge the distance between our participants and ourselves, and to create a more positive attitude towards the assignment. However, we were pleased to also discover that the participants' answers contained both critical and positive remarks concerning their experience. This might be due to the fact that the two researchers conducting the participant recruitment had not been personally involved in the design of the displays. Thus, they were able to tell the participants that they were merely investigating the deployment and the

whole idea of public displays, and that they valued all their opinions. All participants returned the probe notebook, although one had misunderstood the assignment and used an indoor display. The participants were rewarded with movie tickets.

The data gained from the two probes was highly interesting. While it painted a varied picture of experiences and attitudes, several patterns began to emerge already in preliminary analysis. Some of them focused on usability issues, demonstrating that the evaluation probe can certainly help with more open-ended usability evaluation. Many experiences dealt with aesthetic preferences and personal feelings that the participants attached to the displays. They also provided us knowledge about how they viewed this kind of a novel intrusion into their living environment in general. Further, the participants' experiences were found to be strikingly different during the two seasons that posed varied sensory settings [44].

As this was an entirely new application of the probes methodology, needs for further development of the concept did arise. The probes could easily be complemented with group discussions or co-design workshops at the end. After all, the lay expertise developed by our participants would be highly useful in the design of future technologies. Secondly, the probe could be complemented with additional methods, e.g. some preliminary homework or participant video recording; or researchers might conduct traditional semi-structured theme interviews prior to the assignment. This depends largely on research context; what kind of knowledge is sought, and what it will be used for, as illustrated in the Adaptive lighting probe.

### (3) Adaptive lighting probe

The *aim* of this study was to evaluate people's experiences of an adaptive lighting installation in an indoor commercial space. Specifically, the goal was once again to produce rich data on people's experiences for in-depth analyses. The real-world setting presented many challenges from a research point of view. The objective of the new lighting design was to attract customers to the installation area and to invite them to spend more time browsing the products, thus affecting their shopping behavior. The installation was designed and set up in a relatively small and restricted area in the women's clothing section of a department store. The overall illumination of the area was reduced from its surrounding areas. The merchandise and focal points, such as mannequin dolls and background panels, were accentuated by illuminating them with spotlights of static white light (Scenario 1), adaptive white light (Scenario 2), and adaptive coloured and white light (Scenario 3). Overall, the adaptive lighting was meant to serve customers by enabling them to see the products better and enhance their experience. The lighting installation functioned in Oct-Jan, running continuously in a three-day rotation of Scenario 1, Scenario 2 and Scenario 3. The probe study was conducted during the course of approximately one month in Oct-Nov.

All *participants* were females ranging from 35 to 55 years in age since the selected store site contained only women's clothing. The evaluation probe consisted of three visits to the lighting installation site, and a package containing three notebooks and a calendar; however, the probe was also amended with a preliminary semi-structured theme interview lasting approximately 30 minutes. It explored themes relating to commercial spaces, their own shopping behavior, lighting in general, lighting in commercial spaces and the idea of adaptive lighting. The interviews gave us valuable insight into our participants' background and their thoughts about their own shopping behaviour. Several participants noticed the prevalence of lighting related questions. We did not deem it necessary to hide this from our participants; on the contrary, the interviewer told them that they were not test subjects, but partners in the study. The participants were rewarded with cosmetic products of approximately 20 euros in value.



Figure 4. Notebooks in the Adaptive lighting probe package.  
© Anna Luusua 2015

The *physical probe* materials contained three notebooks and a calendar to plan and mark three visits. The notebook format was useful to distinguish the visits from each other, and they were convenient for our participants to carry. A box of items, in the style of cultural probes, would have been a nuisance to carry around. Once again, participants were encouraged to express themselves in any way they preferred; by writing, sketching, or even adding pictures from other sources, e.g. magazines. Each of the notebooks (Fig. 4) was sealed with a sticker, and participants were instructed to break the seals only after having finished the previous notebook. All in all, each participant filled out three notebooks; one for each visit they were required to make in the department store. On each occasion, a different lighting scenario was in operation. The different lighting scenarios, then, were scheduled to rotate on three consecutive days. The participants were given a colour-coded calendar, where each day corresponded with a notebook. The notebooks' colour scheme matched with that of the calendar.

The orchestration of this fairly complex experimental setting proved to be slightly difficult for us and the participants. Some participants actually visited the store on

a wrong day. Further, two holidays occurred within the visitation period, which made the schedule even more complex to understand. Therefore, the practicalities of organizing an evaluation consisting of three visits which necessarily must be done in a certain order are probably too much for participants who are operating independently. In future studies, a comparative evaluation of several scenarios using the evaluation probe method could benefit from a setting which would allow the participants to launch each scenario themselves on site by e.g. a mobile device request.

Despite these difficulties, the probe provided us with hundreds of pages and recorded minutes of experiences and perceptions in the form of sketches, writings, pictures and audio/video. The preliminary interview alone provided us with a rich data set that can be analysed from various points of view: either as background information for helping us in analysing the notebooks, or, for example, as stand-alone material for researching and analysing the participants' attitudes concerning commercial spaces and their lighting. Similarly, with the notebooks, we obtained in-depth accounts of the participants' experiences with the three different lighting scenarios in a real department store context [38].

#### Analysis of evaluation probe data

In all our studies, we used qualitative analysis methods, emphasizing iterative readings of data to understand phenomena. After an initial free-form reading, we structured the materials and our remarks about them into emergent themes [e.g. 6]. These were also organized into matrices to facilitate comparative readings between participants. This thematization and the original data have been analyzed with the aid of theoretical concepts, which have acted as lenses through which we have scrutinized the data. These theoretical lenses include, e.g. *emplacement* [39] and *urban rhythms* [16] in our studies [44, 45]. As previously stated, Gaver *et al.*'s use of cultural probe data was non-analytic. However, we consider the in-depth analysis of the rich evaluation probe data to be central to its utilization and distilling new knowledge.

#### DISCUSSION

In this paper, we have argued for the re-conceptualization of cultural probes towards evaluative research. Through our methodological discussion of probes, we concluded that the original cultural probes contained finely tuned facets that require thoughtful consideration in their application. We argued for epistemological consistency in the design, execution and analysis of probes, especially in multidisciplinary research settings. We accepted functional aesthetics, critical reflection, establishing a rapport and mutual knowledge production, as well as framing the participant as an expert as central features of the methodology.

Our approach combined these with (1) evaluative aims and (2) in-depth theory-based analysis. Thus, the probe consisted of both the evaluated design artefact as well as the

package given to participants. By no means can this new approach be regarded as a cultural probe anymore, and therefore we decided to call it evaluation probes (“EP”). Next we discuss the key benefits and limitations of the evaluation probes discovered in our empirical studies.

#### **Truly “in the wild” with no direct researcher presence**

Firstly, using EP we gained *experiences that were as close to real life as possible* in an emic study, since the users were able to choose the location and the time for using the artefacts in peace, without direct researcher presence, taking their time to hang around, observe the locale and play with the artefact. This, we argue, produces more realistic experiences for participants, which is important in in-the-wild studies with the objective to gain real-world knowledge about design artefacts.

#### **Support for different communication styles**

Secondly, EP *supports different kinds of participants’ communication and learning styles* [12]. This is especially important for in-the-wild studies in which participants experience the research setting with all their senses. In our studies, EP created room for many different kinds of ways to express one’s thoughts: some participants clearly preferred writing; some were more visual and expressed themselves with drawings and clippings; and some made only short notes in the notebook but verbalized their perspectives at length in an interview. However, the probe materials gave everybody a similar opportunity to reflect on their experiences or practices over a longer period of time.

Combining a probe with an interview was found useful, especially regarding the studies’ aim to produce in-depth qualitative knowledge. For example, in the ICT diary probe study, the scrapbooks acted as a preliminary assignment that led to more profound and mature interviews. Further improvements would include the wider inclusion of different sorts of sensory experiences and learning and communication styles through the use of various media. Participants might be given video cameras or audio devices for these purposes. These methods have already been utilised in sensory ethnography [39].

#### **Critical reflection teases out fresh, unexpected views**

Thirdly, EP encouraged the participants to critically reflect upon the artefacts and also their everyday routines. The participants gained a kind of lay expertise on the subject that also challenged the traditional roles of experts and lay persons. For example, while returning the Display probe materials, many participants noted that they now saw the displays differently because they had had a chance to both use them and to reflect upon what the displays meant to them personally. Previously, most participants had not given them as much as a thought in their everyday lives, although they all knew about them. As a result, the participants’ view of the world had changed somewhat as a direct result of the probe study. In the ICT Diary probe, some even reported that they had started to observe other people’s technology use. Everyday life practices tend to

become invisible but the notebook made them visible for the participants. Similarly, the participants of the Adaptive lighting probe noted that they had begun to think of lighting differently over the course of the study.

The thickness and richness of the material itself greatly supports the goal of in-the-wild studies by providing fresh real-world perspectives for designers and researchers. The participants are recruited, and remain, as trusted partners throughout the study. As one possible improvement, the lay expertise created along the study could be later explored further in co-design workshops. As such, EP spans between the information probe and the cultural probe; through understanding participant experiences of the design artefact we can move towards further design ideas rooted in empirical knowledge. We would be especially interested in seeing a process where both cultural probes and evaluation probes would be used at the different stages of a particular design process.

#### **Many types of knowledge from a single EP study**

Fourth, *a single EP study can produce many types of knowledge*. EP typically produces information concerning usability on a practical level, but also knowledge relating to e.g. emotional experiences. The latter is encouraged by the multimodal communication styles supported by EP. As such, EP is an efficient method for conducting open-ended, informal and rich evaluation “in the wild”. However, *probes are also able to produce much more in-depth knowledge that is valuable in and of itself, and, theorized further, can help build and develop empirically-based design theory* on a more fundamental level, as was the case in our own EP studies. The concept can be compared to ethnography – e.g. in our first case these two were merged – as both are qualitative approaches that enable us to map and understand people’s experiences and everyday life.

The outcomes of studies where ethnography and design meet have often been successful [e.g. 8]; nevertheless, Dourish and Bell [13] highlighted how ethnography can produce two kinds of results, both relevant to design: (1) design recommendations, often derived from empirical aspects of ethnographic work, and (2) “profound design guidance” born from the analytical aspects of ethnographical work. The first type of results can provide relatively direct guidelines how to proceed with some limited design challenge; the latter is “where the substantive intellectual achievement can be found”, and it tends to open up new design space rather than closing it down towards ‘the best solution’. In other words, *we call for profound analysis and the use of theoretical lenses also when analyzing probe materials*, as discussed above; although this was not the initial purpose of Gaver *et al.*, it can result in significant insights when applying probes in different ways.

#### **Efficient and flexible evaluation**

Fifth, in our studies we have found that *EP enables researchers and designers to conduct evaluative research*

*efficiently*. If the assignment is carefully designed, the participants are able to proceed fairly independently. EP, then, offers a viable alternative to “rapid ethnography” or “quick-and-dirty ethnography” [35]. Participants are able to spend several days, or even weeks, with the assignment, processing the experience presented to them and exploring their own thoughts. Further, EP is *time flexible*, as demonstrated in the Display probe. While the displays have been deployed in the city for five years, it was especially beneficial that the EP study could be extended for longer or shorter periods of time, and iterated several times over with a reasonable effort. Also, *EP yielded compatible and complementary data to other data sets that we have obtained with other means*. For example, the public displays have generated a wealth of quantitative usage data; the emic, experiential data yielded by the Display probe was able to explain particular phenomena in the quantitative data [44]. Further, our studies show that *EP is able to accommodate different kinds of research settings*. This is especially important in evaluative research, since research settings are as varied as they are numerous.

#### **Meta-participation: participant experiences of probes**

The majority of our participants liked the overall experience and tasks, and they found the colorful and playful design of the probes appealing and motivating. However, some of them thought that the openness of the questions was difficult or confusing, and they wished they had been given more specific questions. This attitude was also present in all our probe participants’ commentary on the concept. Some participants felt that they had already answered a question on a previous page. This was perfectly acceptable, as the questions were meant as a *reflective* framework for the participants. We conducted our readings very freely to foster their personal communication styles. Furthermore, the tasks and questions in the probes were purposefully meant to be open as possible to allow for participants’ own perceptions.

#### **Limitations and caveats**

Although EP produces knowledge through participants’ critical reflection, this also presents a limitation to researchers, as these are not unreflective, humdrum everyday experiences. In other words, there can be no reflection without detachment, which can also be seen as a limitation. The experiences recorded through EP are the experiences of actively thinking participants. Obviously, external observation can be used to observe unknowing passers-by, but their thoughts cannot be documented at all. EP takes us to the limits of what we can learn about experiences, but no further.

Furthermore, the critical reflection of the participants and possible subsequent interchange of roles can also complicate the analysis of the EP data. For example, in the Adaptive lighting probe one participant wrote that she was not at all interested in the clothes, as she was really observing the area and taking notes. Some participants commented on other customers’ behavior at the study site, clearly adopting the

role of an observer. Their perspective is thus moved away from a normal shopping experience, which needs to be acknowledged while analysing the data and interpreting the results, as reflection automatically distances the participants from their daily routines.

In general, the thickness and richness of the EP data can make it relatively hard to analyse. Only the researchers’ imagination limits the different communication modalities than can be included in EP. However, the researchers should carefully anticipate the kinds of data that will be produced and ask themselves whether they will be able to make sense of it. Of course, this kind of freedom of expression also requires that participants are ready to use their creativity and invest time in the study. Rapport-building and enticing materials, then, are absolutely necessary ingredients, making the successful implementation of an EP a delicate balancing act between traditional researcher skills and designer skills.

The use of EP is demanding, as it requires a wide array of skills from researchers, from drafting relevant research questions to designing and producing the probe materials and the research setting, all the way to building rapport with participants and conducting possible interviews and co-design workshops. Thus, it is most useful for heterogeneous but closely knit research and design teams where each individual understands the way in which EP produces knowledge. A foundational understanding of qualitative research and analysis can be considered necessary for conducting an EP study. Similarly, an understanding of the design artifact and the basics of design research is useful.

An EP must be tailored for the participant group that we wish to engage; the participants’ age, gender, and cultural background play an important part in this. Fortunately, the inherent flexibility of EP and the underlying probes approach enables this. The subsequent analysis of the EP data must be conducted accordingly with the tailoring and the participant group in mind.

The researchers’ non-presence also means that no additional questions can be asked while the participant is completing the assignment. Thus, the setting must be carefully considered both in the design and analysis of EP. Overly complex settings are not likely to succeed. Any involved technologies should function at least reasonably reliably to provide the intended experience. If the designed artefact is at its early development stages, or its functionality cannot be guaranteed to a reasonable extent, using EP is probably not the best option. A more suitable qualitative method in these cases might be participant observation coupled with an interview.

However, the lack of researcher presence does not remove the researchers’ accountability and subjectivity from the research process. EP is still designed by the researchers and the results are interpreted by them, which calls for highly reflexive attitude [11, 42]. Further, when using EP as a part

of a design process, the designer's influence within the process must be acknowledged.

### CONCLUSIONS

We presented evaluation probes for conducting emic, experiential evaluation in the wild without researcher presence. We were motivated by the probes methodology being so useful that we deemed it not only acceptable but imperative to develop it for uses outside its original intended domain. However, its use requires open reflection and a highly ethical and critical attitude. Thus, while we are aware of the excited and somewhat haphazard manner in which new types of probes have been developed in design fields, we decided to carefully try and extend it into a new, useful direction. We laid a theoretical foundation for re-conceptualizing and adapting the probe methodology into evaluative in-the-wild practice, basing our approach on a careful analysis of the original cultural probes. Their central features include the functional aesthetics and playful nature of the *physical probe artefacts*; *participation* through building a good rapport with the participants, leading to mutual knowledge production; and the overall *critical and reflexive sensibility* that penetrates the whole methodology. However, we then parted ways with the *aim* of gaining design inspiration of the cultural probes. Throughout the three consecutive studies, we adapted the cultural probe into an evaluation probe that consists of the evaluated design artefact and participant materials, for conducting emic, experiential evaluation of the artefact in-the-wild.

Based on the empirical findings of our studies, we argue that the evaluation probes are highly useful for evaluating urban computing applications from an experiential point of view in the wild. The continued improvement of experiential evaluation methods in urban computing studies is of crucial importance to the well-being and quality of life of urban inhabitants, the number of which is rapidly increasing globally. To introduce computing into the everyday living environment of people is a task that aims to engage the senses, emotions and needs of urban inhabitants on a daily basis. It is the responsibility of the designers and researchers to grasp the consequences of their decisions to the best of their ability. For this purpose, we need methodologies and methods that address these issues.

Finally, evaluation probes are a qualitative research approach which can also be seen as a part of the critical research paradigm, with its accompanying epistemological and ontological groundings. However, through understanding it thoroughly and respecting its qualities, we can also compliment probe materials with other data sets, including quantitative ones. As such, we emphasize the foundational meaning of epistemological consistency and understanding to the larger urban computing research and design community.

### ACKNOWLEDGEMENTS

We thank our participants, the Academy of Finland, TEKES, and the Nokia Foundation for their support.

### REFERENCES

1. Bardzell, J. and Bardzell, S. A great and troubling beauty: cognitive speculation and ubiquitous computing. *Personal and Ubiquitous Computing* 18, 4 (2014), 779-794.
2. Bargas-Avila, J. A. and Hornbæk, K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proc. CHI 2011*, ACM Press (2011), 2689-2698.
3. Bodley, J. H. *Cultural anthropology: Tribes, states, and the global system*. Altamira Press, Lanham, 2011.
4. Boehner, K., Vertesi, J., Sengers, P. and Dourish, P. How HCI interprets the probes. In *Proc. CHI 2007*, ACM Press (2007), 1077-1086.
5. de Bono, E. *Lateral Thinking*. Penguin Books, London, 1990.
6. Bryman, A. *Social research methods*. Oxford University Press, 2012.
7. Chamberlain, A., Crabtree, A., Rodden, T., Jones, M. and Rogers, Y. Research in the wild: understanding 'in the wild' approaches to design and development. In *Proc. DIS 2012*, ACM Press (2012), 795-796.
8. Clarke, A. J. *Design Anthropology: Object Culture in the 21st Century*. Springer, 2011.
9. Cross, N. Designerly ways of knowing. *Design Studies* 3, 4 (1982), 221-227.
10. Danermark, B., ed. *Explaining society: Critical realism in the social sciences*. Routledge, London, 2002.
11. Davies, C. *Reflexive Ethnography: A Guide to Researching Selves and Others*. Routledge, London, 1999.
12. Denig, S. Multiple intelligences and learning styles: Two complementary dimensions. *The Teachers College Record* 106, 1 (2004), 96-111.
13. Dourish, P. and Bell, G. *Divining a digital future: Mess and mythology in ubiquitous computing*. MIT Press, Cambridge, 2011.
14. Downton, P. *Design Research*. RMIT University Press, Melbourne, 2003.
15. Elliott, H. The use of diaries in sociological research on health experience. *Sociological Research Online* 2, 2 (1997). <http://www.socresonline.org.uk/2/2/7.html>
16. Galloway, A. Intimations of everyday life: Ubiquitous computing and the city. *Cultural Studies* 18, 2-3 (2004), 384-408.
17. Gaver, B., Dunne, T. and Pacenti, E. Design: Cultural probes. *Interactions* 6, 1 (1999), 21-29.
18. Gaver, W., Boucher, A., Pennington, S. and Walker, B. 2004. Cultural probes and the value of uncertainty. *Interactions* 11, 5 (2004), 53-56.

19. Graham, C., Cheverst, K. and Rouncefield, M. Whose probe is it anyway? Workshop on Appropriate Methods for Design in Complex and Sensitive Settings at OzCHI 2005.
20. Groat, L. and Wang, D. *Architectural Research Methods*. Wiley, New York, 2002.
21. Harrison, S., Sengers, P. and Tatar, D. Making epistemological trouble: Third-paradigm HCI as successor science. *Interacting with Computers* 23, 5 (2011), 385-392.
22. Harrison, S., Tatar, D. and Sengers, P. The three paradigms of HCI. alt. chi 2007.
23. Hemmings, T., Crabtree, A., Rodden, T., Clarke, K. and Rouchefield, M. Probing the probes. In *Proc. PDC 2002*, ACM Press (2002), 42-50.
24. Hogan, R. Engendered autobiographies: The diary as a feminine form. *Prose Studies: History, Theory, Criticism* 14, 2 (1991), 95-107.
25. Hosio, S., Kostakos, V., Kukka, H., Jurmu, M., Riekkii, J. and Ojala, T. From school food to skate parks in a few clicks: Using public displays to bootstrap civic engagement of the young. In *Proc. Pervasive 2012*, Springer (2012), 425-442.
26. Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, C., Evans, H., Hansen, H., Roussel, N. and Eiderbäck, B. Technology probes: Inspiring design for and with families. In *Proc. CHI 2003*, ACM Press (2003), 17-24.
27. Keskinen, T., Hakulinen, J., Heimonen, T., Turunen, M., Sharma, S., Miettinen, T. and Luhtala, M. Evaluating the experiential user experience of public display applications in the wild. In *Proc. MUM 2013*, ACM Press (2013), Article 7.
28. Kindberg, T., Chalmers, M. and Paulos, E. Guest editors' introduction: Urban computing. *IEEE Pervasive Computing* 6, 3 (2007), 18-20.
29. Kukka, H., Luusua, A., Ylipulli, J., Suopajarvi, T., Kostakos, V. and Ojala, T. From cyberpunk to calm urban computing: Exploring the role of technology in the future cityscape. *Technological Forecasting and Social Change* 84 (2014), 29-42.
30. Kuutti, K. HCI and design - uncomfortable bedfellows? In Binder, T. Löwgren, J. and Malmberg, L. *(Re)searching the Digital Bauhaus*. Springer, London, 2009.
31. Liu, Y., Goncalves, J., Ferreira, D., Xiao, B., Hosio, S. and Kostakos, V. CHI 1994-2013: Mapping two decades of intellectual progress through co-word analysis. In *Proc. CHI 2014*, ACM Press (2014), 3553-3562.
32. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S. and Davies, M. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proc. CHI 2011*, ACM Press, 3033-3042.
33. Mattelmäki, T. *Design probes*. Aalto University, Helsinki, 2006.
34. Mattelmäki, T. and Battarbee, K. Empathy probes. In *Proc. PDC 2002*, ACM Press (2002), 266-271.
35. Millen, D. R. Rapid ethnography: time deepening strategies for HCI field research. In *Proc. DIS 2000*, ACM Press (2000), 280-286.
36. Ojala, T., Kostakos, V., Kukka, H., Heikkinen, T., Linden, T., Jurmu, M., Kruger, F. and Zanni, D. Multipurpose interactive public displays in the wild: Three years later. *Computer* 45, 5 (2012), 42-49.
37. Paulos, E. and Jenkins, T. 2005. Urban probes: encountering our emerging urban atmospheres. In *Proc. CHI 2005*, ACM Press (2005), 341-350.
38. Pihlajaniemi, H., Luusua, A., Markkanen, P., Herneoja, A. and Pentikäinen, V. Experiencing Adaptive Retail Lighting in a Real-World Pilot. In *Proc. Experiencing Light 2014*, Eindhoven University of Technology (2014), 90-93.
39. Pink, S. *Doing Sensory Ethnography*. Sage, London, 2009.
40. Polanyi, M. *The Tacit Dimension*. University of Chicago Press, 1966.
41. Rogers, Y. Interaction design gone wild: striving for wild theory. *Interactions* 18, 4 (2011), 58-62.
42. Schön, D. *The reflective practitioner: How professionals think in action*. Basic books, 1983.
43. Wallace, J., McCarthy, J., Wright, P. and Olivier, P. Making design probes work. In *Proc. CHI 2013*, ACM Press (2013), 3441-3450.
44. Ylipulli J., Luusua A., Kukka H. and Ojala, T. Winter is coming: Introducing climate sensitive urban computing. In *Proc. DIS 2014*, ACM Press (2014), 647-656.
45. Ylipulli, J., Suopajarvi, T., Ojala, T., Kostakos, V. and Kukka, H. Municipal WiFi and interactive displays: Appropriation of new technologies in public urban spaces. *Technological Forecasting and Social Change* 89 (2014), 145-160.
46. Ylipulli, J. and Suopajarvi, T. Contesting ubicomp visions through ICT practices: Power negotiations in the meshwork of a technologised city. *International Communication Gazette* 75, 5-6 (2013), 538-554.