# Prototyping and Sampling Experience to Evaluate Ubiquitous Computing Privacy in the Real World

**Giovanni Iachello[1], Khai N. Truong[2], Gregory D. Abowd[1], Gillian R. Hayes[1], Molly Stevens[3]**

[1] College of Computing & GVU Center
Georgia Inst. of Technology
{giac,abowd,gillian}@cc.gatech.edu

[2] Dept. of Computer Science
University of Toronto
khai@cs.toronto.edu

[3] User Experience Group
Logical Design Solutions
mstevens@lds.com

## ABSTRACT

We developed an inquiry technique, which we called "paratype," based on experience prototyping and event-contingent experience sampling, to survey people in real-life situations about ubiquitous computing (ubicomp) technology. We used this tool to probe the opinions of the conversation partners of users of the Personal Audio Loop, a memory aid that can have a strong impact on their privacy. We present the findings of this study and their implications, specifically the need to broaden public awareness of ubicomp applications and the unfitness of traditional data protection guidelines for tackling the privacy issues of many ubicomp applications. We also point out benefits and methodological issues of paratypes and discuss why they are particularly fit for studying certain classes of mobile and ubicomp applications.

### Author Keywords

Experience Sampling, Paratype, Privacy, Ubiquitous Computing, Mobile Computing, User Centered Design.

### ACM Classification Keywords

D.2.2 [**Design Tools and Techniques**]: Evolutionary Prototyping; K.4.2 [**Computers and Society**]: Social Issues; K.8.m [**Personal Computing**]: Miscellaneous.

## INTRODUCTION

Over the past several years, we have developed and evaluated several ubiquitous computing (ubicomp) applications, including many based on mobile platforms such as cell phones. These applications present numerous challenges to effective user-centered design (UCD). First, understanding usage environments requires designers to step out of the lab and follow people where they use these applications—on streets, in shopping malls, homes and wherever else they might go. Second, evaluation must also occur in a "situated" setting in order to account for physical and social interactions, disruptions, variations in cognitive load, and other environmental factors that can profoundly affect the usability and usefulness of mobile applications [1].

Work in the HCI community has started to focus on in-the-field prototyping to address these challenges: the Experience Prototypes used by Buchenau and Suri [6] and Wizard-of-Oz techniques used in mobile settings [29, 11] are aimed at evaluating mobile technologies within iterative development processes. Unfortunately, these approaches base their evaluation on reproduced or simulated experiences with the aim of testing the technology, instead of *measuring the experience* with reference to specific instances of real life.

This lack of reference to actual experience represents a serious methodological flaw when probing some social domains. Privacy represents one such domain. Researchers have long recognized that people often take a deontological stance when artificially probed on opinions and preferences on privacy, both in reference to organizations [5] and in interpersonal relations. Everyday behavior may differ from stated preferences for many reasons, including insufficient informational awareness (ignoring the fate of collected information), overriding primary goals (getting a transaction done), or carelessness (not wanting to bother with evaluating every exchange of information), as pointed out by Acquisti and Großklags [3]. On the other hand, people have a very refined sense of privacy balance in interpersonal relations, as described by Altman [4], and may choose certain paths of behavior to avoid conflict or in response to overriding social goals.

It follows that abstract or purely self-reflective surveys may be insufficient for probing privacy concerns. This observation applies in general to all those situations in which people may be unable or unwilling to explain their behavior abstractedly, from complex social constructions, as suggested by Goffman [18], to the formulation of procedural plans as noted by Suchman [35]. In addition, people may be unable to grasp immediately the effects of new technologies on their existing socio-technical practices.


*CHI 2006*, April 22–27, 2006, Montréal, Québec, Canada.

This article offers two contributions. First, we present a user inquiry tool that we used to probe people's opinions on the privacy impact of an ubicomp technology with reference to real-life situations. Second, we present the results of this study, which may have a broad impact on the debate on privacy in ubiquitous computing. Below, we discuss the evaluation challenge that prompted us to develop this instrument. Then, we describe the inquiry method, and present the results and their consequences on the design for privacy. We finally reflect on the advantages and drawbacks of this inquiry technique.

### The Evaluation Challenge

Over the past two years, we have worked on a personal, portable, audio memory aid called the Personal Audio Loop, or PAL [20]. PAL was motivated by everyday conversational breakdowns experienced when people try to remember something that was just said—the topic of a conversation before an interruption, or a name or number heard in situations of high cognitive load such as at a conference or a party. In the initial concept, the user would be able to replay, at any moment in time, sounds heard in the recent past, up to a defined maximum time span (for example, up to one hour in the past). Audio older than this *retention time* is automatically deleted. Figure 1 shows how PAL was implemented on the Motorola i730 phone. PAL records whenever the phone is in the closed position and the user is not using the phone for telecommunication. The recording can be replayed by the press of a button.

Initial interviews and a diary study showed that participants would find this tool very useful in many situations. They reported that they would use it a few times a week for various purposes, the top three being to remember forgotten details of conversations, replaying conversations to their conversation partners, and recovering the topic of a conversation after an interruption.

Although study participants recognized the usefulness of PAL, they also raised concerns about this technology. PAL runs continuously and unattended, and unnoticed by the user and potential conversation partners or bystanders. The primary concerns related to the impact on the privacy of secondary stakeholders (conversation partners) and unrelated third parties (*e.g.*, passersby). Further concerns related to the social appropriateness of using the application, regarding both the immediate disruption of interpersonal interaction and long-term effect on social relationships.

After one short deployment of PAL, some participants reported on the reactions of secondary stakeholders to their using the application, noting that in many cases conversation partners would not object to the use of PAL after being briefed on its purpose and characteristics (*i.e.*, the limited retention time and the impossibility of permanently storing the recording). However, they also said that they avoided mentioning the presence of PAL in some situations, reportedly to avoid explaining its features time and again. In some


**Figure 1 The Personal Audio Loop. Main picture: the device and its interface. Three buttons control record/playback mode and allow to browse the recording. A timeline and a mode light indicate the current operating state. Upper-Right picture: the device can be attached to a belt clip.**

settings, they spontaneously turned it off. Participants seemingly adopted self-regulating boundary-setting towards their conversation partners' privacy, similar to what was described by Palen and Dourish [34].

Far from providing conclusive answers, these anecdotal observations raised further issues regarding contextualization and adjudication. The question of contextualization is whether and how the social and technical environment affects the use of the application. If different privacy requirements could be associated to specific circumstances (*e.g.*, when driving alone), and at least part of these circumstances could be detected by a recognizable "context,"[1] this could bring about significant design implications, as suggested by the work of Hong and Landay [22].

The issue of adjudication raises two questions. The first question relates to whose interests should prevail. The primary stakeholder (the user) of the application may have a legitimate interest in using PAL, for example due to a memory dysfunction or simply because of cognitive stress imposed by his or her occupation. This interest may be opposed to that of secondary stakeholders or third parties (who might not want to be recorded, even if only temporarily). The second adjudication question relates to the proportion of individuals opposed to the application. If only a small minority of secondary stakeholders and third parties oppose PAL and the vast majority does not care, should we

[1] The term 'context' is used in the ubicomp community to indicate information about the user's social and technical environment that can be used to drive the operation of IT applications (*e.g.*, the user's location, people present) [10].

yield to the contrary minority and curtail a large market potential?

This is a classic issue in the ethical debate between utilitarianism and normativism in the context of privacy and technology [37]. Designers might not be in the position of deciding on these issues—they rest in the hands of courts and Data Protection Authorities (DPA), with the final judge being market acceptance. However, designers should still consider these issues, at least to increase profitability and hedge liability. In a previous article at this conference, we suggested performing these determinations borrowing judgment tools from the legal community [25]. In this case, to make any such determination we needed to understand the following:

1. *to what degree*, and *in what situations* secondary stakeholders are most likely to object to the use of a device that can potentially cause the recording of their conversation (*i.e.*, are objections unqualified or do they depend on the location, on the topic, on the identity of the conversation partner, or on the perceived confidentiality of the conversation?); and
2. *what application parameters* (*e.g.*, retention time) can be adjusted to meet a compromise between the interest of the primary users and conversation partners.

Limitations of the prototypes prevented us from investigating these questions through a long-term deployment that may have provided strong qualitative and quantitative evidence. However, we were wary of using a privacy survey, abstract from social practice, for the reasons discussed above. So, we designed an event-contingent experience sampling procedure targeted at the conversation partners of PAL's user. Wheeler and Rois define event-contingent sampling procedures as those initiated by the occurrence of a specific event, in this case a conversation [41].

## METHOD

The study was conducted by a group of individuals whom we called "proxies" as they acted as substitutes or probes for the researchers. The proxies were two females and one male between the ages of 27 and 31, all HCI graduate students and researchers working at this institution, living alone or with their partner. As the proxies went about their daily activity, they asked people older than 18, acquaintances as well as unknowns, at the end of their normal conversations, to read a description of PAL and to fill out an anonymous survey about the application. The proxies were instructed to hand out the surveys only if the conversation exceeded 3–4 sentences—weeding out very short interactions such as "thank you" and "excuse me." Also, we only submitted the survey once to the same individual to avoid respondent bias. The survey was administered in various locations in the United States.

### Using Paratypes

The survey asked participants to suppose that the researcher has been using PAL, and probed their opinions and feelings about the operation of PAL in *that specific conversation*. In this respect, this technique is similar to Critical Incident techniques developed in the context of workplace psychology [13]. This procedure allowed us to situate participant response in the experience the person just had, with a specific partner, conversation topic and location, supposedly reducing recall errors and hypothetical answers. We call this experimental procedure a *paratype*: a simulation, or model, of interaction ("-type") with a technology which is evaluated alongside ("para-") real-world experience. That the survey was administered by human proxies is not part of our definition of paratype. The term paratype only refers to introducing simulated interaction with a certain technological artifact within a specific setting of real social action, and documenting the effects of this combination. The proxy's role was only incidentally that of administering the survey—her main function was that of acting as "PAL's user" and as interaction counterpart of the participant. In this sense, the proxy's role was to *create the technological instance* on which we wanted feedback, with the help of the description of the application and, if requested, a demonstration of the working device. Event-contingent experience sampling was deemed a particularly suitable way of documenting participant feedback in this case.

The survey was composed of two parts, linked by a unique number (Figure 2). The first part (on the left) was filled out by the investigator, with information about place, participants, and the activity being achieved with the conversation, as salient elements of the social setting [17]. The portion on the right was detached and given to the participant after the conversation between the proxy and the participant had occurred. The participant was also given an IRB (Institutional Review Board)[2] consent notice at this point.

The survey was designed to be self-explanatory, and contained a description of PAL and a short questionnaire. The description had been validated prior to the study to ensure that it would be pertinent and sufficient. When possible, the investigator explained PAL verbally, and optionally showed the working application, if requested by the participant. We chose not to operate PAL during the study on ethical grounds and to avoid contentious situations.

The participant was asked to fill out the survey immediately if possible, to increase recall accuracy. Otherwise, the survey portion of the card (lower right in Figure 2) was return-addressed on the backside and could be mailed back at the participant's convenience (we affixed a postage stamp for

[2] The Institutional Review Board is an organization, present in most US research institutions, responsible of protecting the rights and welfare of human subjects involved in research, as mandated by State and Federal Government.

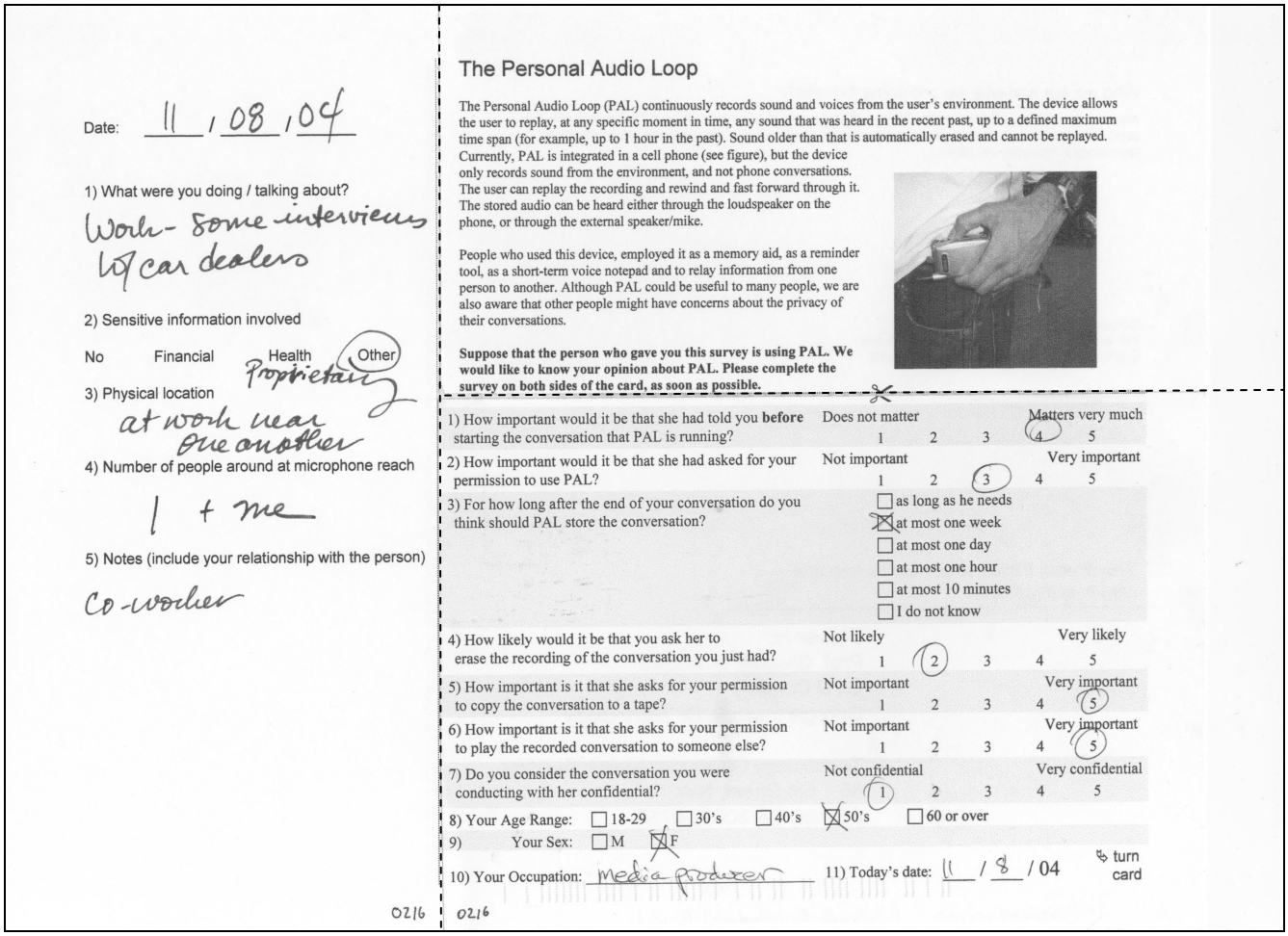


**Figure 2 The survey form is divided in three parts, here shown after being reassembled. The left side is filled out by the researcher. The right side is given to the participant after the conversation. It contains a description of the application and the survey. The backside of the participant's portion is addressed to the researchers' lab, like a postcard.**

this purpose). The questionnaire included six questions on a 5-point scale, one multiple-choice question and, on the backside, a blank space for optional comments in addition to our lab's address and space for postage. The questions included the following:

- the importance of being informed about the application;
- the importance of asking permission before using the application;
- the time span for which the subject would allow the user to store the conversation;
- the likelihood that the subject would ask the user to erase the recording;
- the importance of asking for permission to copy and replay the conversation to others; and
- an indication of the subjective "confidentiality" of the conversation.

The survey also included three anonymous demographic questions: age range (in decade), gender and occupation. This structure minimized completion time and, in fact, most participants were able to complete the survey immediately.

## RESULTS

Of 45 distributed surveys, we received 41 usable responses. This represents a very high response rate, possibly attributable to the personal contact the participant established during the conversation. Only one person refused to accept the survey at all. Most surveys were completed immediately, and 9 were mailed back to us afterwards.

### Demographics

24 respondents were in IT or research occupations (students, research scientists, university professors, *etc.*). The remaining respondents ranged across professions, including: teachers, designers, hairdressers, managers, attorneys and business owners. Respondents spanned all age groups between 18 and 60 and over. However, age distribution was biased towards the younger age groups (the median age group was 30–39), reflecting the age group of the proxies. 17 respondents were female.

We did not observe strong correlation between the opinions expressed by participants and type of occupations (correlation with technical occupation is low for all questions on the survey). We also did not observe strong influence of age and gender. We did not plan to analyze the impact of the relationship between the proxy's and the respondents' genders or ages.

### Survey Responses

Both the proxy and the participant were asked to provide a measure of the sensitivity or confidentiality of the conversation they had just conducted. The participant was asked to rate subjective "*confidentiality*" on a 5 option scale. Proxies indicated whether the conversation was *sensitive* following precise guidelines given by data protection legislation [12 §8] (*i.e.*, financial, health, religious and some work-related topics are sensitive).[3] We would like to stress the difference between confidentiality and sensitivity: we did not observe significant correlation between sensitivity and confidentiality as indicated by the proxies and the participants (see below for a discussion).

#### *Informed Consent*

Participants wanted to be informed that the recording is happening: 30 participants responded 4 or 5 on the respective question, with avg. 3.8, $\sigma = 1.4$, where 1 is "not important" and 5 is "very important" (Figure 3). Participants wanted to be informed regardless of the sensitivity of the topic of the conversation as classified by the researcher and the place where it happened (public or not). There was, however, a positive correlation between perceived confidentiality and this variable ($r = 0.58$; $p < 0.001$).

Participants indicated that it was important that the PAL user (the proxy) ask for permission before using PAL in the conversation that had just occurred (avg. 3.8, $\sigma = 1.2$, where 1 is "not important" and 5 is "very important.") Again, there was positive correlation between perceived confidentiality and this variable ($r = 0.60$; $p < 0.001$).

#### *Further Use*

Participants stated it was important that the person using PAL ask permission before he or she copies (avg. 4.0, $\sigma = 1.2$, where 1 is "not important" and 5 is "very important") or replays (avg. 4.3, $\sigma = 0.9$, where 1 is "not important" and 5 is "very important") the audio to others. There were relatively weak correlations of these two variables with confidentiality ($r = 0.40$ and $r = 0.36$ respectively for copying and replaying to others, both $p < 0.05$). This suggests that there is a concern with what happens with the recording, regardless of its perceived confidentiality. The interpretation is that copying the conversation onto other media or replaying it to others may cause the de-situation of the interpersonal communication experience and could engender uncomfortable situations or social breakdowns.

#### *Deletion and Retention*

The desire to be informed does not necessarily imply that the participants would have likely asked to erase the recording after the fact. In most situations people would not have asked to erase the conversation (avg. 2.3, $\sigma = 1.2$, where 1 is "not likely" and 5 is "very likely"). There appears to be a "confidentiality threshold" in this respect: participants would have asked the user to erase the recording primarily in cases of elevated confidentiality. In most cases of conversations of medium or low confidentiality, they would not have likely asked to delete the recording. This is supported by the very strong correlation between confidentiality and likelihood of requesting deletion ($r = 0.70$; $p < 10^{-6}$). In fact, only 7 responses indicated 4 or 5 on this question, supporting anecdotal evidence gathered during the deployment and comments arising from formative evaluation [20].

Finally, participants indicated that a long retention time would not be an issue (Figure 4): 43% of respondents indicated that the user could have kept the conversation for "as

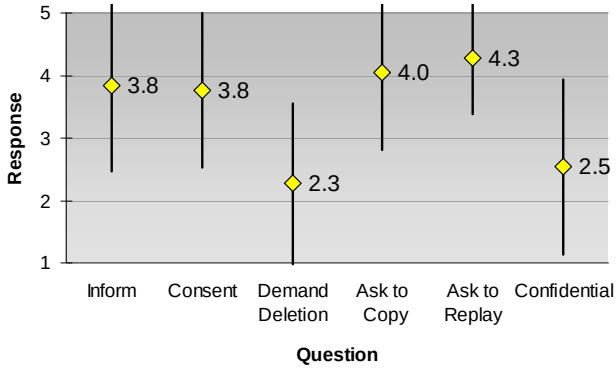


**Figure 3 Results of the survey. The diamonds indicate the average response to each question. The bars indicate standard deviation.**

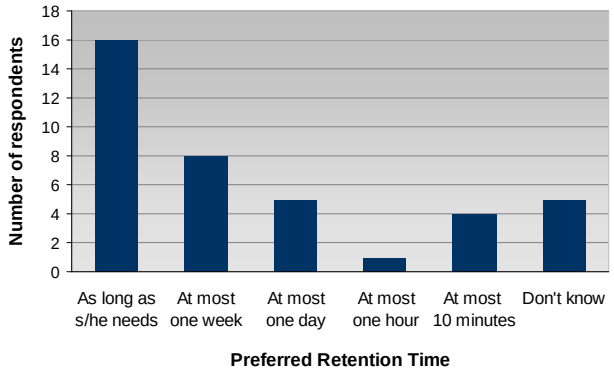


**Figure 4 Preferred Retention Time. Two participants did not respond to the question. Both indicated that they would very likely ask to delete the recording.**

[3] Proxies did not rate confidentiality because we wanted them to record information as objectively as possible.

long as he/she needs to" and 22% "at most one week". Our original estimate of appropriate retention time was in the range of 10 minutes–1 hour. That estimate attempted to minimize privacy risks while retaining application usefulness and usability, but these participants were comfortable with a much longer retention time. Apparently participants were more concerned with the *misuse* of the recording (*e.g.* by replaying to others) than with its mere *storage*.

## PRIVACY IMPLICATIONS

When Weiser spelled out the vision of ubicomp, he suggested that mature technologies fade out of the user's consciousness [40]. Likewise, informed consent dynamics will also have to become implicit. Most interviewed participants desired being informed about the application when it was running and being able to consent to its use. The tension here is that explicit informed consent may be so cumbersome as to make PAL unusable in its current format. However, public knowledge and awareness of technology is an integral part of this changing picture. Even a technology as disruptive from the privacy standpoint as camera phones has been widely adopted within the relatively short span of two years. Today the public has evolved compensatory behaviors and norms [36, 39], to make carrying camera phones in places such as public restrooms acceptable (as long as they are not used). Informed consent procedures have become *implicit*, because the public is learning how camera phones are used and misused.

This observation suggests that designers and researchers should devote more attention to the adoption of ubicomp technologies. Norman has written about the adoption of ubicomp (or 'information appliances') in *The Invisible Computer*, characterizing users in terms of their place in the adoption curve [33]. We suggest designers focus on predicting and influencing adoption patterns *as part of the design process*. Similar to how Christensen *et al.* have attempted to predict adoption phenomena to drive investment decisions [7], we believe that ubicomp design would benefit from incorporating an adoption strategy. Feeding knowledge and hypotheses about long-term adoption patterns into the UCD of ubicomp applications would allow designers to better define design hypotheses (*e.g.* in terms of legality or social acceptability) and foresee possible critics. This knowledge might improve our ability to influence public awareness and to identify emergent concerns, lessening the cultural stress that many indicate today as an acceptance challenge of ubicomp and facilitating the process of "shifting the technology to the background."

Participants expressed a desire to be informed and asked permission to use PAL. Yet, in many cases, they reported not being likely to ask the user to erase the recording after the fact. This assessment is hypothetical in nature and should be taken with caution. If the situation probed with the questionnaire would have happened with PAL actually operating, outside of a study setting, the conversation partner might have been even less inclined to make such a request. This apparent contradiction may be explained by social dynamics that induce people to overlook incidents to avoid conflict or uncomfortable situations (notwithstanding a potential longer-term impact).

Participants also stated that they would allow the user of PAL to retain their conversations for much longer than we had expected. Data retention time has served as a fundamental principle for the data protection community for years, and this has influenced the ubicomp security field. Jiang *et al.*, in their Approximate Information Flows framework, modulate information "persistency" to achieve privacy goals [26]. Unlimited data retention has also been identified by Palen and Dourish as one problem with pervasive information collection and aggregation [34]. These participants did not express their concerns and needs in terms of retention time; the focus was rather on misuse and social appropriateness. This may be due to the fact that people are unable to express appropriate retention policies without reliable information on risk of misuse over time. This suggests that both in talking with users and in designing ubicomp applications, analysts should focus on the *purpose of use* of the information rather than on its *temporal retention*.

Above we reported a lack of significant correlation between the determination that the conversation's topic was sensitive according to data protection guidelines and the confidentiality measure indicated by the participant. The guidelines used to categorize sensitive conversations may not accurately reflect interpersonal dynamics. The definition of sensitive information we used was based on the data protection community's description (*i.e.*, financial, health, religious, and some work-related topics [12 §8]). This definition may not be descriptive of the knowledge that is likely to be relevant in interpersonal dynamics, such as social "secrets" [18] or other topics of conversation that can significantly affect social relations if misused.

Furthermore, classic data protection guidelines are difficult to apply to many small collections of personal information (*e.g.*, recordings of conversations taken for personal purposes), as opposed to large organizational data banks. For example, the costly requirements imposed by legislation on these data banks (informed consent, redress, *etc.*) are inappropriate to personal collections of data. One important conclusion is that the issues involved in the negotiation of privacy between individuals are different from those related to data protection. If this is true, design guidelines [14, 27] inspired by data protection principles or the Fair Information Practices (FIPS) [38] may not provide sufficient guidance for the design of technologies that increase the information sharing between individuals.

We believe that, in general, privacy-enhanced design of these ubicomp technologies may be better served by shifting the focus from algorithmic approaches (*e.g.*, defining

"personal privacy policies" upfront, or modeling preferences quantitatively in terms of retention time or a reliability percentage) to the dynamics of social interaction, at least when talking with users and when designing user interfaces.

## BENEFITS AND CAVEATS OF PARATYPES

We believe that the experimental setup described above allowed us to gather observations that would have been difficult to obtain with a typical survey or in a laboratory setup. For example, we could have used scenarios to evaluate PAL, asking people to read a short story of one instance of use of PAL and then answer to a survey. Such a setup would have three drawbacks: 1) it would require participants to reflect on a situation that may not be part of their way of behaving; 2) the selection of the situation and the way the scenario is presented may affect the results; and 3) it would be difficult to create a set of scenarios representative of all possible conversational situations.

Because paratypes combine event-contingent experience sampling with experience prototyping, they are especially useful for evaluating high-level or implicit interaction where reference to concrete instances of life is needed. This is particularly important when we do not know how often these instances arise, or how to describe them (*e.g.*, how often are people in a "confidential" conversation?) and when later recollection of these instances may be inaccurate. The latter may occur when reference is sought to social relations, privacy, or personal preferences, or when considering situated action [35].

We claim that the procedure that we called paratype can be particularly useful for gathering early feedback on mobile and ubiquitous technologies, such as applications that collect information at unexpected times (*e.g.*, Microsoft's SenseCam [16]), that provide information when needed (*e.g.*, portable guides), or where interaction is embedded in unplanned social practice or everyday routine, such as home communication systems [31]. These applications may have high prototyping costs; probing salient aspects of the experience without the need of working artifacts may represent an efficient way of obtaining relevant information needed for design. However, when using this kind of inquiry techniques, several issues must be considered, including potential sources of bias, study implementation cost and IRB compliance.

### Sources of Bias

The demographics of the proxies are likely to influence the demographics of the respondents, in terms of age, socioeconomic class, education, *etc.* In this study, the proxies' age, profession and social class are reflected in the respondents'. Also, relative differences in age, gender and socioeconomic class between proxy and respondent could influence the results. To control for these variables, it is advisable to recruit a diverse group of proxies as possible, and to verify that their social interaction patterns actually reach the target demographic by using exercises specifically developed for this purpose [30].

Selection bias on the probed situations may also represent a potential issue. The protocol called for submitting the survey systematically to all qualifying interactions for a short period (some days). However, one proxy felt it was impractical to distribute the survey in some situations, for example, when trying on clothes in a shop or one time when she was in a hurry at the grocery store. Moreover, the proxy did feel it was inappropriate to hand out surveys in some situations, or to some people, such as at a medical visit—social settings which are arguably very interesting from the standpoint of the research questions we investigated in this study. The need for probing these situations depends on the intended context of use of the technology under investigation. Research ethics and social appropriateness may suggest avoiding intrusive studies in sensitive situations (*e.g.*, at a funeral) and to resort to more hypothetical means for probing behavior.

However, some parameters might be evaluated by extension. In the PAL case, participants desired to be informed about the presence of the device, independent of the topic of the conversation or the place where it takes place—there is little reason why such a desire should not also apply to the delicate situations described above. Clearly, the responses on deletion and retention cannot be generalized in the same way, being functions of the conversation's perceived confidentiality.

Certain countermeasures may decrease or control situation selection bias. First, the researcher administering the survey could complete his side of the data, without handing out the survey to the conversation partner. This would provide a statistic on the number of situations that were not probed, and the reason why this did not happen, thus allowing researchers to plan supplemental inquiry. Second, administration of the experience sampling survey could be deferred to a more appropriate moment. Researcher's notes could be used to help the participant recollect the situation, although there is a risk of introducing even more bias.

Ultimately, it may be difficult to justify serious infringement of social norms just for the sake of accurate behavioral research for product development. This limitation should be worked around or accepted in many cases (but not always: see Hudson and Bruckman's study on chat room privacy where some norms were infringed [24]). Unfortunately, we did not record statistics on situation selection bias by the proxies. However, informal feedback from proxies suggests that our results portray a reasonably accurate generalization of the target population's reactions within the type of (normal) situations probed. We would strongly encourage anyone utilizing these inquiry techniques in the future to adopt at least one of the above two countermeasures to control this source of bias.

### Cost

Administering a survey is decidedly less expensive than a deployment with a working prototype, which typically has high development, recruitment, and operational costs. In fact, one of the reasons we chose not to deploy PAL in a long-term study was that we were unwilling to commit the resources required for product "hardening" and maintenance of the deployment. However, this survey is not, by any means, a discount usability technique [32]—it still required careful planning and its execution was more complex than a mass survey administered via email or by stopping people in a shopping mall.

Sample size is related to cost. The number of responses might seem low compared to other privacy surveys such as the GVU WWW User Surveys [19] and Ackerman *et al.*'s e-commerce surveys [2]. This number derives from the structure of the study. We wanted to investigate initial reactions to the technology in real-life situations and we avoided submitting the survey twice to the same person. The number of respondents was thus bounded by the number of significant social interactions with unique individuals that the proxies had within the timeframe devoted to the study (limited by cost considerations). This number is limited by the size of the *interactive network* of the proxy and is surprisingly low for many people, as studies of social relationships and frequency of communication indicate [21, 30]. In effect, we traded quantity for increased authenticity and *situatedness*. Demographic coverage may represent an issue if it does not match the intended population and must be controlled. However, with appropriate coverage, a small number of respondents is not necessarily a limitation. Even with 41 responses, we were able to produce relatively good statistics (in terms of *p* values). Higher collection costs per response (*i.e.*, fewer responses) are the price to pay for presumably more accurate and salient responses.

### Survey Administration and IRB Protocol Compliance

The potential problems arising from the application of IRB requirements to social, behavioral and economic research are well known (see, for example, the National Research Council's report of 2003 on the topic [8]).

One practical problem we incurred was related to consent requirements set by our IRB. Although we did not have to document participants' consent (*i.e.*, have them sign a consent form), we still had to provide them with an information notice. Reading the one-and-a-half page notice disrupted the experience even further than the disruption caused by filling out the survey. More concise consent notices would be helpful, though changing standard wording requires extensive collaboration with IRB officials.

Our IRB also ruled that if the proxy was initiating the conversation, he or she had to inform the participant upfront about the survey to avoid the possibility that the participant might think the conversation was initiated solely for the purpose of the survey (which would imply a form of deception). Fulfilling this requirement proved to be very cumbersome and we do not see how it could be avoided, except by filing a protocol with participant deception.

## RELATED WORK

HCI practitioners and researchers have long worked to probe the user's experience of a not yet existing artifact or technology. Prototypes have often been used for this purpose: Houde and Hill point out that prototypes can reproduce the user interface of products, but also serve to probe their functional and experiential characteristics [23]. Buchenau and Suri describe the use of Experience Prototypes to evaluate the user experience of new products such as a digital camera [6]. They point out that role playing can be useful for understanding the social context in which a person will use a technology. Our "paratype" can be viewed as a form of experience prototyping. The difference between Buchanau and Suri's prototypes and our procedure is that they recreate and study simulated experiences (*e.g.*, by building a reproduction of a plane interior, or by role playing during a train journey). Our procedure leverages real-life situations. The conversations on which we seek feedback are real, as they would have happened independently of the survey.

In this sense, paratypes are *situated* in the real world, and explore interpersonal interactions, not unlike Cultural Probes. Cultural Probes introduce foreign artifacts (post cards, maps, *etc.*) in participants' life and use them to document and communicate about everyday experience and social practices in a situated manner [15]. Paratypes introduce a foreign technology in real-life situations to understand the social practices that would arise from the presence of such technology.

Psychologists have long asked people to report on their experience in specific circumstances with the purpose of uncovering behavioral patterns. Flanagan performed pioneering work analyzing statistical data on critical incidents to evaluate, among other things, individuals' proficiency at performing work tasks and reasons for success and failure [13]. This Critical Incident Technique was also used for analyzing human-machine interaction, using incident reports collected after the fact to improve aircraft instrument design. The inquiry technique presented here attempts to generate a more accurate reproduction of stakeholders' interests by eliminating recall and selection bias.

More recently, Wheeler and Rois examined of the use of self-reported inquiry techniques within behaviorism, medicine, and industrial psychology [41]. They categorize these techniques in: interval-, signal- and event-contingent, depending on what initiates the self-report procedure. Larson and Csikszentmihalyi were among the first to propose experience sampling as a quantitative self-reported inquiry technique [28] for social and psychological research. We used a form of event-contingent experience sampling, fo-

cused on the participants' activities and feelings about the technology introduced by the paratype.

Experience Sampling is becoming increasingly popular in HCI practice, often in association with diary studies. Consolvo *et al.* conducted a signal-contingent study of the responses to simulated location requests, focusing on privacy issues [9]. Participants carried a Palm device that simulated at random times location requests from friends, family and colleagues. Consolvo *et al.* point out that the random simulated requests that the device made to participants were in various occasions implausible from a social standpoint. For example, the device would simulate a person asking for the participant's location when the participant felt that the person would have not done so in reality. The event-contingent nature of our study makes the responses presumably more salient.

Our paratype was based on a paper survey. An interesting question would be if more sophisticated setups would be feasible, such as providing realistic prototypes to support the experience, either functioning or through Wizard-of-Oz. In addition to the ethical questions discussed above, it is necessary to consider whether increasing prototype realism is necessary to probe user opinions, and whether it is cost-effective to do so. The nature of ubicomp technology may help in this case, because PAL operates in the background and is invisible to the conversation partner, so no real artifact is needed. Furthermore, Wizard-of-Oz setups may be difficult to apply to ubicomp technologies in real-world settings. Although it is true that Wizard-of-Oz techniques have been used in the past to test mobile technologies (*e.g.*, in the Topiary system by Li *et al.* [29] or in the use of the DART toolkit for mixed-reality tour guides [11]), it is also true that these setups require fabricated usage situations akin lab experiments due to the need for a wizard to be present. They may thus be unfit to the unpredictable situations probed by the paratype.

## CONCLUSIONS

To understand the impact on privacy of a personal memory aid, we developed an inquiry protocol based on event-contingent experience sampling and experience prototyping that we called "paratype." Paratypes probe a technological experience in the real-life context where it happens. We claim that paratypes may be particularly fit for studying a large class of mobile and ubicomp technologies where naturalistic authenticity is required and prototyping costs are high. However, their use requires solving practical roadblocks related to the disruption in the flow of social interaction, sampling accuracy, and procedural requirements.

The study provided interesting results on participants' opinions on privacy. Participants stated that awareness about PAL was important to allow "boundary-setting" to occur. They were not concerned as much by retention time as with potential misuse of the recordings. They also stated that they would have rarely asked to delete a recording after the fact. These observations have broad consequences because they suggest that traditional privacy guidelines and quantitative privacy policies may not be appropriate or sufficient for the development of this kind of personal ubicomp applications and that designers should focus on the purpose of use of information and interpersonal dynamics instead.

## ACKNOWLEDGMENTS

We thank Shwetak Patel for writing the PAL application, Jens Großklags, Elaine Huang, Kris Nagel, Delphine Nain and Jay Summet and the CHI reviewers for their thoughtful comments and all study participants. This work was supported in part by NSF ITR grant 0121661, the NSF GRFP, the MacArthur Foundation, Intel Corporation and Motorola Corporation. This work is covered under IRB protocol H04153 at the Georgia Institute of Technology.

## REFERENCES

1. Abowd G.D., Mynatt E.D. Charting Past, Present and Future Research in Ubiquitous Computing. *ACM ToCHI 7*, 1 (2000), 29–58.
2. Ackerman, M., Cranor, L., Reagle, J. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. *Proc. E-Commerce 99*, ACM Press (1999), 1–8.
3. Acquisti, A., Großklags, J. Privacy and Rationality in Individual Decision Making, *IEEE Security and Privacy 3*, 1 (2005), 26–33.
4. Altman, I. *The Environment and Social Behavior—Privacy, Personal Space, Territory, Crowding*. Brooks/Cole Publishing Company, Monterey, CA, 1975. ISBN 0818501685.
5. Berendt, B., Günther, O., Spiekermann, S. Privacy in e-commerce: stated preferences vs. actual behavior. *Commun. ACM 48*, 4 (2005), 101–106.
6. Buchenau, M., Suri, J.F. Experience Prototyping. *Proc. DIS 2000*, ACM Press (2000), 424–433.
7. Christensen, C.M., Anthony, S.D. Roth, E.A. *Seeing What's Next: Using Theories of Innovation to Predict Industry Change*, Harvard Business School Press: Boston, MA, USA, 2004, p.xi+312. ISBN 1591391857.
8. Citro, C.F., Iglen D.R., Marrett C.B. (eds.) *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. National Academies Press, 2003.
9. Consolvo, S., Smith, I., Matthews, T., LaMarca, A., Tabert, J., Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share. *Proc. CHI 2005*, ACM Press (2005), 82–90.
10. Dey, A.K., Salber D., Abowd G.D. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction Journal 16*, 2–4 (2001), 97–166.

11. Dow S., Lee, J., Oezbek, C., MacIntyre, B., Bolter J.D, Gandy, M. Exploring Spatial Narratives and Mixed Reality Experiences in Oakland Cemetery. *Proc. ACE'05*, ACM Press (2005).
12. European Parliament, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *European Union Official Journal L281* (1995), 31–50.
13. Flanagan, J.C. The Critical Incident Technique. *Psychological Bulletin 51*, 4 (1954), 327–359.
14. Garfinkel, S. Adopting Fair Information Practices to Low Cost RFID Systems. *Proc. Ubicomp 2002 Privacy Workshop* (2002).
15. Gaver, B., Dunne, T., Pacenti, E. Cultural probes. *Interactions: New Visions of Human-Computer Interaction 6*, 1, ACM Press (1999), 21–29.
16. Gemmell, J., Williams, L., Wood, K., Lueder, R., Bell, G. Passive Capture and Ensuing Issues for a Personal Lifetime Store. *Proc. CARPE'04*, ACM Press (2004), 48–55.
17. Goffman, E. *Behavior in Public Places*. Free Press, 1966. ISBN 0029119405.
18. Goffman, E. *The Presentation of Self in Everyday Life*. Anchor Books, 1959. ISBN 0385094027.
19. GVU Center, *10th WWW User Survey – Online Privacy and Security*, 1999. http://www.gvu.gatech.edu/user_surveys/survey-1998-10/graphs/graphs.html#privacy.
20. Hayes, G.R., Patel, S.N., Truong, K.N., Iachello, G., Kientz, J.A., Farmer, R. Abowd, G.D. The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid. *Proc. Mobile HCI 2004*, LNCS 3160, Springer Verlag (2004), 168–179.
21. Hill, R.A., Dunbar, R.I.M. Social Network Size In Humans. *Human Nature 14*, 1 (2003), 53–72.
22. Hong, J., Landay, J. An Architecture for Privacy-Sensitive Ubiquitous Computing. *Proc. MobiSys 2004*, ACM Press (2004), 177–189.
23. Houde, S., Hill, C. "What do prototypes prototype?" *Handbook of Human-Computer Interaction (2nd Ed.)*, Helander M., Landauer T., Prabhu P. (eds.). Elsevier Science B. V. Amsterdam, 1997.
24. Hudson J.M., Bruckman, A. Using empirical data to reason about Internet research ethics. *Proc. ECSCW '05*, Springer Verlag (2005), ISBN 1-4020-4022-9.
25. Iachello, G., Abowd, G.D. Privacy and Proportionality: Adapting Legal Evaluation Techniques to Inform Design in Ubiquitous Computing. *Proc. CHI 2005*, ACM Press (2005), 91–100.
26. Jiang, X., Hong, J.I., Landay, J.A. Approximate Information Flows: Socially-Based Modeling of Privacy in Ubiquitous Computing. *Proc. Ubicomp 2002*, LNCS 2498, Springer Verlag (2002), 176–193.
27. Langheinrich, M. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. *Proc. Ubicomp 2001*, LNCS 2201, Springer Verlag (2001), 273–291.
28. Larson, R., Csikszentmihalyi, M. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science 15* (1983), 41–56.
29. Li, Y., Hong, J., Landay, J. Topiary: A Tool for Prototyping Location-Enhanced Applications. *Proc. UIST 2004*, ACM Press (2004), 217–226.
30. Milardo, R. M. Comparative methods for delineating social networks. *Journal of Social and Personal Relationships 9* (1992), 447–461.
31. Nagel, K., Hudson, J., Abowd, G.D. Predictors of availability in home life context-mediated communication, *Proc. CSCW'04*, ACM Press (2004), 497–506.
32. Nielsen, J. *Usability Engineering*. Academic Press, Boston, MA, 1993, p.xiv+358. ISBN 0125184050.
33. Norman, D. *The Invisible Computer*. MIT Press, Cambridge, MA, USA, 1999, p.320. ISBN 0262640414.
34. Palen, L., Dourish, P. Unpacking "Privacy" for a Networked World. *Proc. CHI 2003*, ACM Press (2003), 129–136.
35. Suchman, L.A. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press (1987) ISBN 0521337399.
36. Sung-jin, K. Camera Phone to Require Shutter Sound From Next Yr. *Korea Times*, 11/11/2003.
37. Terrell, T., Jacobs A. Privacy, technology, and terrorism: Bartnicki, Kyllo, and the normative struggle behind competing claims to solitude and security. *Emory Law Journal 51*, 4 (2002), 1469–1511.
38. United States Department of Health, Education and Welfare. *Records, Computers and the Rights of Citizens, Report of the Secretary's Advisory Committee on Automated Personal Data Systems*. 1973.
39. United States Video Voyeurism Prevention Act of 2004. *18 USC §1801* et seq.
40. Weiser, M. Some Computer Science Problems in Ubiquitous Computing. *Commun. ACM 36*, 7 ACM Press (1993), 75–84.
41. Wheeler, L., Rois, H.T. Self-Recording of Everyday Life Events: Origins, Types, and Uses. *Journal of Personality 59*, 3 (1991), 339–355.