Improving Consent in Large Scale Mobile HCI through Personalised Representations of Data

Alistair Morrison
School of Computing Science
University of Glasgow, UK
alistair.morrison@glasgow.ac.uk

Donald McMillan
Mobile Life Centre
Stockholm University, Sweden
don@mobilelifecentre.org

Matthew Chalmers
School of Computing Science
University of Glasgow, UK
matthew.chalmers@glasgow.ac.uk

ABSTRACT

In using 'app store'-style software repositories to distribute research applications, substantial ethical challenge exists in gaining informed consent from potential participants. Standard 'terms and conditions' pages are commonly used, but we find they fail to communicate relevant information to users. We suggest interrupting use of an application with a visual representation of collected data, rather than merely providing a description at first launch. Data collected, but not uploaded, before this can be used to create personalised examples of what will be shared. We experiment with different ways of presenting this information and allowing opt-out mechanisms, finding that users are more concerned when presented with a visual, personalised representation, and consequently stop using the application sooner. We observe a particular difference in non-English speakers, suggesting that our proposed approach might be especially appropriate for global trials, where not all users will be able to understand researchers' disclosures of data logging intent.

Author Keywords

Ethics; informed consent; app stores; large scale trials.

ACM Classification Keywords

H.5.2 User Interfaces: Evaluation/methodology

INTRODUCTION

Large scale mobile HCI trials using 'app stores' as a recruitment platform are becoming popular [5]. Such trials benefit from the potential to reach tens of thousands of users without the practical challenge of supplying hardware to such a large population. While reaping these benefits, researchers should be aware of a set of ethical challenges arising from conducting research in this way [25]. One significant consideration is in the gathering of informed consent from trial participants. Accepted ethical practice in human trials dictates that each participant should be instructed on the nature of the experiment before taking part, and told about the data to be collected on their behaviour or use of trial software. This could be seen as particularly necessary in large scale mobile HCI trials, where it is possible for users to be continually monitored via their mobile devices over a long time period, yielding vast amounts of personal information. A consensus is still lacking as to how researchers involved in such research should best inform users and gain consent.

Here we describe exploratory work on gaining more reliably informed consent from users, beyond potentially ineffective standard 'terms and conditions' (T&C) pages or end-user license agreements (EULAs). Following strategies used to increase the end user understanding in Web and desktop applications [11, 30], we experiment with delayed intervention, where the information is presented to interrupt use of the app while participants are engaged. We take an exploratory approach to studying issues surrounding T&Cs and informed consent, examining different ways of presenting relevant information including a visual, personalised representation of recorded data. In light of this, we analyse users' reported level of comfort about being logged, as well as effects on continued use of the application. We also look at the effects of language differences on these issues, and finally, we provide greater control over the data recorded, in order to see which types of data cause users the most concern.

RELATED WORK

Human Subject Research and Informed Consent

Any field of research that uses human subjects is normally guided by an ethical code of practice which sets out the responsibilities researchers have towards safeguarding their participants. As we discuss in previous work [25], both the British Psychological Society (BPS) and the American Psychological Association (APA) provide principles-based guidelines for researchers conducting human trials and both of these sets of guidelines stress the participants' fundamental rights to self-determination, autonomy and personal liberty – that participants have to be briefed on the nature of an experiment, and that participation must be optional. Such rights are usually safeguarded by the gathering of *informed consent* – an (often signed)

NordiCHI '14, October 26 - 30 2014, Helsinki, Finland

ACM 978-1-4503-2542-4/14/10…$15.00.
http://dx.doi.org/10.1145/2639189.2639239

declaration by a participant that they understand the consequences of participation in the experiment. These consequences could include various forms of data being recorded on them, and the sharing or publishing of this information.

Informed consent has been adopted as a standard practice in many fields of research, being seen as fulfilling ethical responsibilities towards, among other areas, privacy and data protection [9]. Governance of such research in academia typically comes via approval by an institution's ethics committee or Institutional Review Board (IRB).

In more traditional HCI software trials, participants are often recruited locally, through posters or mailing lists and come to a specific location to run a study under researchers' observation or to have software or hardware distributed to them directly. Even away from the lab in fieldwork such as ethnographic studies of a workplace, it is common for researchers to meet participants being studied, and there is an opportunity to discuss the work and the participant's role in it. Crucially, during this meeting there is usually the possibility for the researchers to assure themselves that the participant understands the consequences of the study. This can be done in advance, and again via a debriefing session on completion of the experiment if necessary.

A move beyond these methodologies, to less controlled environments, began when researchers started performing research into online behaviours and communities. For example, there has been discussion over ethnographic research of chat rooms, and arguments over whether this might be exempt from requiring informed consent [15].

Whatever the consensus might be on this point, we believe that a separate decision has to be made for the kinds of smartphone data logging that is the subject of this paper. This form of study cannot simply be thought of as a subset of other forms of distance or online research due to the amount of data and the many different forms of personally identifiable information that may be captured in a background, 'always on' manner.

Large-Scale Mobile HCI through 'App Stores'

Following the popularity of app stores on all major mobile platforms, several academic research projects have begun to use such stores as a means of recruiting participants and deploying software [5]. The benefits of this methodology are numerous. As massive numbers of users can be relatively easily recruited [26], it allows researchers greater certainty in drawing conclusions based on consistently observed behaviour. It also allows for greater geographical and, potentially, cultural diversity among participants, enabling a better understanding of whether results generalise across such contexts.

Several projects have looked at different types of questions using this recruitment method, such as assessing touch performance [14], analysing installation and usage patterns [2], and examining the trial methodology itself [26, 28].

Challenges in Gaining Informed Consent in Large-Scale Mobile HCI

There are known to be challenges in operating ethically when conducting large scale trials of mobile HCI [25]. Researchers will generally not physically meet participants recruited through an app store, and will not be able to verbally explain the nature of a trial. At a distance, it becomes difficult for researchers to assure themselves that the user understands the consequences of participation and is consciously agreeing to take part.

A further drawback is the inability to debrief participants following the trial. A user can stop using an application at any time (long before actively uninstalling the software) and researchers will be unlikely to have any forewarning of this disengagement with the application. As the primary means of communication with users is likely to be through the app itself, there may be no opportunity to pass such debrief information to participants after the point at which use of the application ceases.

A workaround among researchers that has perhaps become the de facto standard [13, 26, 31] is to follow the example of commercial software and provide boilerplate T&C screens. These screens will often declare that the software is part of academic research, and might explicitly require users to agree that they have read and understood the terms before they can continue to use the application.

For many apps, T&Cs are shown *only* on first launch [13, 28]. This is most likely done with good intentions –users must give consent before any potential data logging begins – but it also means that if a user doesn't take the time to read them at first launch, he or she might never take the opportunity to learn about the study and any available opt-out mechanisms, even if the information is available within the application or supporting website.

Despite the recent relative popularity of large scale mobile HCI trials, few studies have discussed ethical issues, or reported their efforts to ensure consent in any detail. An experiment by Pielot et al. [31] showed that subtle differences in the presentation of consent dialogs can affect the proportions of users accepting the terms, although the authors note that they do not address whether any of the tested methods are ethically sound. Henze et al. [13] report on several large scale mobile HCI studies they have performed. In a section on ethical considerations, the authors state that their mobile apps do not access personally identifiable information such as unique hardware identifiers, and that location is not recorded at the highest possible accuracy. The authors feel that this absolves them of the need to gain explicit user consent for performing other forms of data logging, and state that in some apps the user is not even informed that logging is occurring.

A 'hybrid' method discussed by Morrison et al. [28] acknowledges the problems of gathering informed consent at a distance and suggests a means to help address ethical concerns by running a large scale mobile HCI trial alongside a more traditional local trial. Different forms of analysis could then be undertaken with the two user groups, with more invasive or potentially sensitive data only gathered from the local users, from whom there was more certainty about the gathered consent.

To foreground the implications of data logging Khovanskaya et al. [17] experimented with showing back data to users in provocative but relevant ways in order to raise awareness and user autonomy.

Public Concern with Regards to Privacy & Data Sharing

In 2000, a poll of attitudes to privacy online found that 56% of people would always opt out of having their personal information collected if they could, while a further 34% would sometimes opt out [3]. User attitudes might have shifted or softened since then in response to technological developments or the dissolution of traditional social barriers of privacy, driven in part by the poor privacy controls commonly provided by online social networking sites [32]. More recent research suggests that the majority of users are generally relaxed about having activity recorded or sharing personal data, but that a minority of users exist who report great concern [18, 34].

However, even if there has been a recent shift in user attitude, it cannot be expected to be consistent across, or even within, cultures and demographics. Given the amount of data and scope for personal identification of users, we would argue that researchers should adopt a cautious approach.

Lawmakers have recently begun to respond to these issues, with the recently passed EU directive on the movement of personal data [4], and the U.S. Federal Trade Commission issuing a report in February 2013 with advice for mobile app developers and researchers as well as guidance for the mobile platforms themselves [7]. The same body also fined Path, a two-year-old social networking app, $800,000 (a record high for an app developer) for violating federal privacy protections for children by collecting personal information on underage users [33]. These developments should re-emphasise that researchers should always do all they can to ensure their studies abide by the law.

Terms & Conditions and End User Licence Agreements

A survey study into the attention paid to T&Cs during installation of desktop software found that only 1% of users always thoroughly read EULAs, 25% regularly browse them but 74% rarely or never even browse the contents [11]. Another study talks of users dismissing installation-time EULAs “so quickly that they did not even remember clicking through the short notices and the Microsoft warnings as they popped up” [10]. Egelman [6] similarly found that consent failures were most likely to occur when users didn’t notice the relevant information rather than a result of them not understanding.

Böhme and Köpsell performed a large-scale field study of over 80,000 users and showed that users become more likely to ‘blindly’ accept terms the more similarly they resemble a EULA—a situation they describe as “thwarting the very intention of informed consent” [1]. Recently the contents of the EULAs has been shown to be too complex for the purpose of informing users [21], further reinforcing the requirement to look for alternative means of informing participants and gaining consent.

Strategies for Better Informing via EULAs

Friedman et al. [9] discuss ways to design for informed consent and present six principles to guide the design of consent interfaces, focusing largely on disclosure and targeting user comprehension. Much of this discussion concerns Web browser interfaces and the work pre-dates app stores and the methods of data gathering that are the focus of this paper.

Patrick and Kenny tried to solve the issue of long, unread agreements by proposing a series of shorter agreements spread throughout software usage, known as “Just-In-time Click-Through Agreements” (JITCTAs) [30]. This approach replaces T&C agreement on first launch with smaller JITCTAs, displayed when a user activates the specific software functions that lead to the access of personal information. The rationale behind this is that users will be more likely to read and understand smaller forms of information, shown at more appropriate times than first launch. However, one of Friedman et al.’s principles concerns “minimal distraction”, and cautions against asking for specific consent from users too often, as this might lead to ‘interruption fatigue’.

Good et al. have looked at the timing of T&C presentation and users’ feelings of regret after installing potentially privacy-breaching software, advising that “it may even be worth preventing immediate use of software to provide a period of reflection” [11]. This work also advocates the use of shorter summary notices and suggests that more personalised notices might be more effective than the more common ‘one size fits all’ approach.

Kay and Terry have aimed to present more visual EULAs that adopt strategies used in other visual media to draw attention, and which highlight information of personal relevance. Their approaches were found to greatly increase the amount of time users spent reading the agreements, and participants’ retention of the information presented [16].

On a higher level, Luger & Rodden [20] suggest that informed consent should be more than simply a notification, and that methods for achieving it should maintain awareness of and reconnect users with their data

The methods we suggest in this paper draw inspiration from these strategies – showing information other than long T&Cs, at times other than at first launch, and of a more personalised nature.

USER STUDY

In the following sections of this paper we first seek to establish whether the current common practice of using

T&Cs pages as a means of gathering informed consent in downloaded mobile apps is effective, or whether this practice suffers from the same problems as reported in studies of desktop EULAs.

Secondly, we experiment with an alternative, or an addition, to first-launch T&Cs, where we interrupt participants during their use of software to remind them that they are participating in a trial. As this intervention would occur after some use of the application, it is possible to show users some personal data recorded from their own usage. We explore different presentations of this logged information including different levels of personalisation and examine how this affects users' level of concern about data logging, and their continued use of the software. As these are global trials, and users will have many different native languages, we also study whether there are differences in the reactions of English and non-English speakers. Finally, we explore whether users have blanket levels of comfort about being monitored in general, or whether there are particular types of data about which they are especially concerned.

Hungry Yoshi

In order to run this study, an existing mobile software application was adapted. Hungry Yoshi is a mobile game on the iOS platform [26]. The game is based on Wi-Fi infrastructure in players' environments—each non-password protected access point is designated as a plantation growing a certain type of fruit, and each password protected access points is a creature ('a Yoshi'). The goal of the player is to pick fruit from the plantations and carry it to one of the creatures that wants that type of fruit. The game was originally studied qualitatively to see how users would integrate gameplay into their everyday lives. Figure 1 shows screenshots from the game.

The iOS app was released publicly in 2009. Log data is gathered on user actions, such as button taps and screen changes. We also record environmental and contextual data, such as location and Wi-Fi access points detected. This information is written to local text files on the device and periodically it is securely uploaded to our servers using the same network connection required to play the game. More qualitative data is gathered by built in questionnaires, where completion is rewarded with in game points.

Figure 1. The Hungry Yoshi app on iOS showing (L-R) a list of detectable access points, a Yoshi, and the logging control panel added for this study.

From its first public release, the application has had a T&Cs page. This page explains the nature of the research and who is conducting it, providing a URL to the project website. It explains all forms of logging, informs users of opt-out procedures and assures users that we will destroy all data recorded on them should they contact us at the provided email address. As we cannot be certain that all users will speak English, all this information is presented in four languages. These terms are shown to users on first launch after the application is installed, and have to be agreed to before use of the game can commence.

User Numbers and Demographics

For the purposes of this study we adapted the app to present new questionnaires and to record different forms of data. The study of this paper is split into three sections. The first part, which looks at effectiveness of standard T&C pages, was run based on the original release and considers results from 75,818 users. The second part, looking at new methods for interrupting users to show them personal data, was based on the updated version of the app, which was used by 1,007 participants. A further update was released for the final part of the study, and looks at 881 users. Due to the nature of such large-scale research, and the incremental release of updates to the software, these three parts can be considered to use independent sample groups, with little overlap of users.

It has been noted that providing demographic figures for this sort of study is difficult [29]. For example, we cannot determine user gender other than via a questionnaire, and the users who provide an answer to this might not answer our other research questions, or vice versa. However, with these caveats, of the users who registered, 11% reported their gender: 65% as male and 35% female. The app was released on an APT-based repository for jailbroken iOS devices. We have previously discussed demographic issues between this repository and the official App Store [24]. It is unclear whether or how this might bias the sample of users recruited; users might choose to jailbreak because they are less risk-averse than a 'standard' user, or might do so to gain greater control of the device and to install further privacy-protecting services.

For the second part of our study, users were split into A-B groups based on hardware identifiers. Each device has a globally unique identifier, or GUID. A hash on the GUID can be calculated and used to assign each user to a group, giving a roughly 50-50 split.

DO PARTICIPANTS OF MOBILE STUDIES READ THE TERMS AND CONDITIONS TO WHICH THEY AGREE?

As previously explained, the app provides a T&Cs screen on first launch after download. All users must agree to these T&Cs before use of the game and before data collection begins.

One of the questions we put to users in the in-app questionnaire was whether they had read these T&Cs and understood that their use of the game constituted participation in an academic trial. Of 1,226 responses, only 20% stated that they were aware of this.

Further, another section of the questionnaire asked for volunteers to participate in a telephone interview about the software. We conducted interviews with 11 volunteers. All these users were asked the same questions about T&Cs and their understanding about the trial, and all 11 stated that they had not realised they were participating in an academic trial, and had not properly read the T&Cs.

As well as directly asking users, we can log data to further investigate this question. The software was modified to put the full T&Cs information on a separate screen in the app, so that we could record the number of people who would open this view, and the length of time they spent on this page. The log data recorded tells a similar story to the questionnaire and interviews: of the 75,818 people who agreed to the T&Cs, only 2% opened the full document. Of these users who did open it, not one spent longer than 60 seconds reading the document, which was 842 words long in English, and similar lengths in the other languages, meaning that it is unreasonable to expect that much of what is contained in this document has been read and understood even by these users [23].

From these findings, it appears that our mobile app users do show similar behaviour to those reported in the literature in studies of desktop EULAs. We therefore suggest that the common practice of using T&Cs pages in large scale mobile HCI trials is not sufficient to gather informed consent. To behave in an ethical manner, in a way that supports users' autonomy, we must seek further means of informing users of our intentions with regards to data capture and potential invasions of their privacy.

DISCLOSING LOGGING BY INTERRUPTING USE, FORMS OF PRESENTATION & EFFECTS ON CONCERN

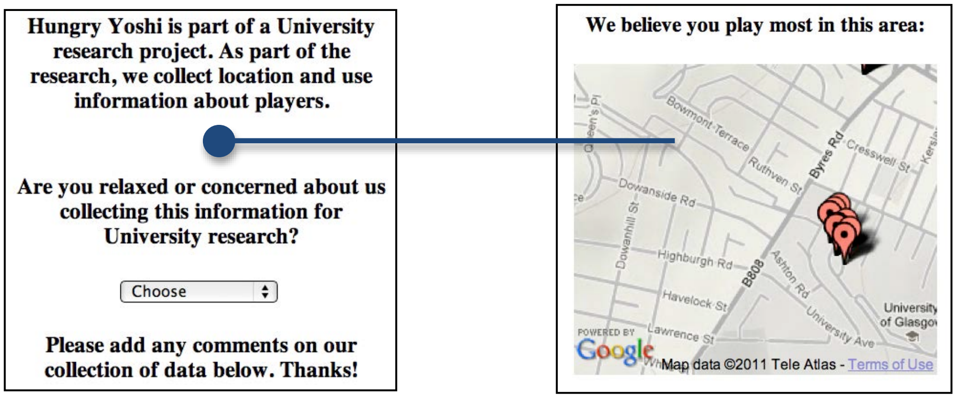
Having established that users are not reading T&Cs, we now therefore look to the use of additional methods of revealing to users the data that we are recording on them. Following work on JITCTAs and the timing of T&Cs [11, 30], we experiment with showing T&C-style information not only on first launch, but interrupting users to present the information again at a later stage. As this will occur at a time when data has already been collected, the T&C information can contain a personal representation of the user's data.

We test two different forms of information presentation: textual descriptions and visual presentations of recorded data. This allows comparison between a standardised presentation where every participant is shown the same information, and a more tailored presentation where the displayed data is personal to the user to whom it is shown.

To explore this, we used a between-groups experiment design. As described above, hardware identifiers were used to assign users to groups. In this study, one experiment group and one control group were created, with a roughly 50-50 split between the 1007 participants. Each group was exposed to different information.

We interrupted the users with this presentation once they had uploaded enough locations that our system could generate a 'home' map tile for them. On average, participants were interrupted after having played the game for 4.3 days (this figure being calculated as the number of unique days the player launched the application). The longest number of days somebody played before being presented with the information was 24 days. The shortest number of days was 1 – 16% of users had provided enough usage data to be interrupted on their first day of use. The control group was interrupted with the same timing – when the user had generated sufficient data that would have qualified them for the experiment group.

Both groups were first reminded that the game recorded information that could identify their locations. The control group was shown the text on the left of Figure 2. For the experimental group, each user's recorded location data was clustered to identify the places he or she had used the application most frequently, along with the text "We believe you play most in this area". These points were shown on a map, as in the example on the right of Figure 2.



Hungry Yoshi is part of a University research project. As part of the research, we collect location and use information about players.

Are you relaxed or concerned about us collecting this information for University research?

Please add any comments on our collection of data below. Thanks!

We believe you play most in this area:

Figure 2. Survey question shown to users. Both groups were shown the text (left), but users in the experimental group were also shown the map (right). The user's location data is clustered, and the map indicates the recorded regions at which the user has most frequently used the app. The control group was shown no such user–specific recorded data.

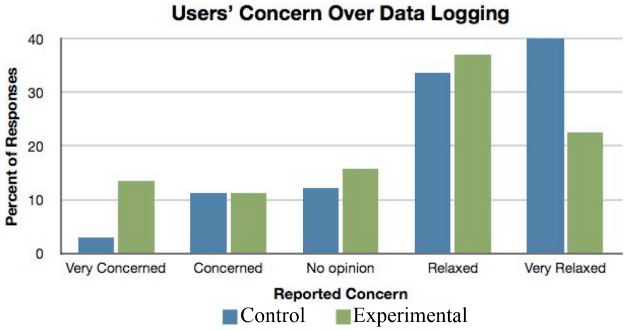


Figure 3. Reported concern about application logging by the experimental and control group members.

Members of both groups were then asked to rate their comfort with the data recording on a 5-point Likert scale ('Very Concerned' to 'Very Relaxed'), and a text box was provided to optionally provide a more detailed response.

Figure 3 shows the results of the Likert scale test. Comparing the distribution of the two conditions, it can be seen that the only large differences between the two conditions occur at the extreme ends of scale. 10% more users who were shown the map claim to be very concerned with being logged than the control group (13% as compared to 3%), and the number of very relaxed users differs by 17%: 40% (control) compared to 23% (experimental). More formally, we performed a Mann-Whitney U test to analyse the hypothesis that the distributions of the two experimental conditions shown in Figure 3 are different. We found a significant difference between the two groups ($p<0.01$), with more concern reported if the map is shown.

We can also look at qualitative data on this issue. In addition to providing a level of happiness, 22% of users also submitted text. Again, most responses were provided by relaxed users. Comments received included "I have no problem! if this help to gather info I would participate in it" or "I hope that your project goes well". Yet a few quotes expressed serious concern, such as "I don't like that people knows that much about me" or "U should not hav the right to do this!!!it's invading in peoples privacy :(now can u pls stop doing this!!".

Effect of Concern on Subsequent Use

As explained earlier, the main drawback of requiring users to agree to T&Cs as a means of gathering informed consent is that users will tick a box without having understood or even read the information being presented to them. The experiment of the previous section was designed to explore an alternative to the T&Cs approach, but our experiment into user attitudes could be viewed as also suffering from the same drawbacks as T&Cs themselves: the users could again have quickly ticked one of the 'concern level' boxes at random, in order to clear the interruption to their game.

Therefore, in order to provide further verification of these results, we compare the submitted answers to information that can be gathered from the log data. We examine the link between the concern the user expressed and the number of days the user continued to play after giving this response. This allows us to see the extent to which the comfort survey results are reflected in observed behaviour. The user sample for this question was the same as the previous section. As stated, participants had been playing for an average of 4.3 days before being presented with the survey.

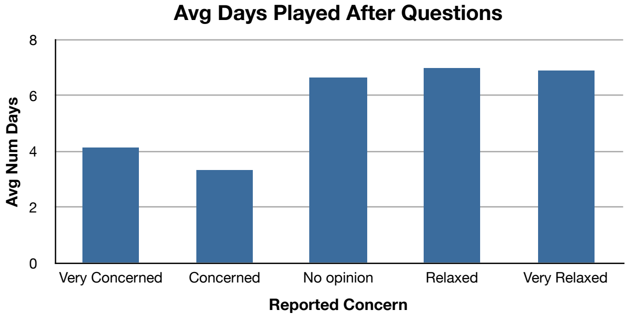


Figure 4. The average number of days the users continued to use the application after being asked about logging.

Figure 4 shows the results of this comparison. It splits users by their reported level of concern and shows the average number of days each group continued to use the application after having been presented with the survey question. We note that users who offered responses of Concerned (C) or Very Concerned (VC) stopped playing considerably sooner on average. These results do not fall on a completely consistent scale, with VC users continuing to play longer than C, and Very Relaxed (VR) users stopping slightly sooner than Relaxed (R). However, averaging the groups who were relaxed (R and VR) and the groups who were concerned (C and VC), we see that the average length of continued play for concerned users was 3.7 days, with relaxed users continuing to play for an average of 6.9 days.

Comparing English and non-English Speakers

The next factor under investigation is whether there is any difference between English and non-English speakers in the days played following interruption.

iOS allows detection of the language the user has set for the device. The most common language among our users was English, being set as the OS language on 64% of the devices from which we had gathered log data. The second most common language was Spanish, on 11% of devices. Unlike gender, this information can be gathered from the device directly and therefore is more complete and reliable data, although we note that we cannot tell from these measurements the level to which the 'non-English speaking' users understood English – only that their preferred OS language was not English.

We found a difference between English and non-English speakers in the number of days played following presentation of the map. For the non-English speakers, the experimental group (shown textual and map presentation) stopped playing after an average of 5.6 days following presentation of the data, whereas the control group (shown text only) continued playing for 8.9 days – a difference of

	Control	*Exp*	*Difference*	*Significance*
Eng speaker	5.25	4.35	21%	p < 0.01
Non-Eng	8.91	5.57	60%	p < 0.0001

Table 1. Average number of days played following presentation of data, split by experimental group and language settings of device.

60% longer. For the English speakers, the difference between the maps vs. text-only groups was only 21%. Using a Welch's t test, the effect of presentation for English speakers was significant with $p<0.01$ and for non-English speakers the effect was significant with $p<0.0001$. This information is summarised in Table 1.

Gender and age differences

Previous works have suggested that concern might be affected by other factors, such as gender or age [22]. Therefore, as well as language, we also compared results across other demographic variables. We looked at gender, age (both obtained from questionnaire answers) and device type – whether the game was being played on an iPhone, iPod Touch or iPad (obtained from the log data). However, in these comparisons we found no significant differences.

USERS' ACCEPTANCE OF TYPES OF LOGGING

As Figure 4 shows, even those users who expressed serious concern about the logging process continued to use the application for some time after the intervention, on average continuing for another 4 days. This is somewhat counterintuitive, and suggests that although these users were unhappy about having their usage recorded, some were willing to continue in return for use of the game.

It may be the case that the greater one's engagement with an application the greater the concern felt when a perceived breach of trust takes place. It might also be the case that these users were concerned, but continued to use the app cautiously. For example, if users were concerned about having their home location logged, they may have decided to only use the software in public places.

In order to explore this further, the final part of our investigation explores users' attitudes to different types of data logging. We examine whether there are some forms of data users are less comfortable about us recording, and look to identify which specific types of data are considered to be most concerning. Further, we wish to see whether there are some types of information users would be willing to let us record, while disabling the logging of other forms of data.

The log data we collected can be classified into three main areas: location, device information (unique identifiers, hardware models etc.) and game-specific information such as button taps. For the purposes of this study, the app was updated to provide users with finer control over data logging. We added a control panel for all users of the application which explained the forms of logging and provided three separate logging control switches. These allowed a user to stop any or all of these forms of data recording, as shown in Figure 1 (right).

	Location	*Device*	*Game*	*All forms*
All users (100%)	17%	15%	15%	14%
Changers (19%)	87%	77%	77%	72%

Table 2. Percentage of users who, given the choice, chose to turn off each form of data logging. Row 1 looks at the data from all 881 users; row 2 only considers the 122 users who made any form of change to the settings.

This adaptation to the software allowed users to continue to use the game while withholding some or all data, opening a dialogue between users and researchers as suggested in [35]. Prior to these changes, had they been concerned about the data logging, their only course of action was to cease playing the game. We sought to explore the number of users who would wish to continue playing but prefer not to be logged on certain aspects of their usage.

Of the 881 participants in this part of the study, 81% chose to keep all logging, with only 19% making any alterations.

Only considering those users who made any alterations, we observed that 87% turn off location logging, 77% turn off game-specific logging and 77% turn off device information logging. A large majority of those who made any alterations – 72% – turned off all logging. Location being the most frequently turned off is in line with work showing its higher perceived value and risk in relation to other types of personal data [19]. This is summarised in Table 2.

DISCUSSION

We began by noting findings on read rates of desktop EULAs, and testing whether a similar pattern would be observed with the T&Cs pages commonly used in mobile HCI trials. The results from our questionnaires, interviews, and quantitative analysis of log data all confirm this hypothesis; not enough attention is being paid to our current T&Cs pages for their continued use to be a reliable solution to gaining informed consent.

Questioning levels of concern among the users of a mobile app, the majority seem to be relaxed about data gathering, with only 17% of users overall reporting that they were concerned or very concerned. These numbers back up more recent findings in the literature. However, we found significant difference in expressed concern when users were shown personalised visual representations of the data rather than impersonal textual information. An interruption containing only a textual disclosure of information at this point, though unexpected compared to up-front T&Cs, might just be another piece of text to click through and ignore. Users claim to be more relaxed when seeing the textual report that we are logging data, but they perhaps do not fully understand all that this entails; only by also seeing the map do they understand exactly what information we have recorded about them, and are therefore in a better position to judge whether they wish to continue to participate. We propose that showing visual representations of this information results in users who are better informed.

The data recorded on days played following the survey backs up the reported concern, yet we do note the possibility that those who said they were concerned might have been answering in a manner they felt was expected. For example, a question on security or privacy might encourage concern.

In reflecting log data back to participants in this manner, we observe that some users are unhappy about the data we are recording and consequently stop using the software more quickly. In effect, our revelations of data logging have 'scared' them away. Yet, this is not an absolutely clear-cut process. We see from Figure 4 that Very Concerned users continue to play for longer than Concerned, and that Very Relaxed users cease play sooner than Relaxed.

To explore further, we added tools with which users could limit the capture of certain types of data, and to investigate whether there are specific forms of information gathering about which users are particularly concerned. We suggest that our techniques presented above are more solid ethical practice, as we have stopped the collection of data from those who were not willing to give consent to the experiment. Yet many might not view this as an ideal solution, as we are reducing the amount of data captured – possibly to a level below the threshold at which the experiment becomes unviable. Allowing users to opt out of certain forms of data logging might be an acceptable compromise; they may be happy to provide some forms of data, and therefore still usefully contribute to our research, as long as they have the ability to limit our capture of other types of information that they are more reluctant to share.

After providing users with finer controls over what types of data were logged and looking at which aspects of logging caused users most concern, it was notable that most users who changed their logging settings chose to switch off all three types of logging at the same time. While we would wish to accommodate user requests if there were *some* aspects of logging they were specifically concerned about, this lack of discrimination in logging is discouraging. Researchers are in effect compensating users for their participation, not with money but by providing a free application, despite associated maintenance and running costs. There is little research benefit in continuing to allow to use the system if they have turned off all logging.

We believe it would be generally acceptable to limit access to the application to those willing to provide data to the study for which it was developed. The one caveat to this would be in cases where the application developed was social in nature and pervasive of a social group, in which case the social pressure to participate may infringe on the autonomy of the users.

The global nature of such trials also raises the issue of language differences. If users cannot comprehensively understand T&Cs presented in English, non-English speakers might be considered particularly vulnerable to having their autonomy impeded when deciding on the information they are providing to researchers. Our results showed that the differences in days played between the map and text-only conditions were even more pronounced for non-English language speakers. The non-English speakers were especially affected by a personalised graphical presentation as compared to a block of text that they perhaps could not read. This suggests that visual disclosures of the data collected are a more appropriate means of informing users in a global release.

Generally, our results suggest that the practice of showing the visual displays is worthwhile in supporting user autonomy. We obviously only want to capture data from users who are happy to provide it, and the users who express great concern about data logging are probably not people we want to involve in our trials. If, through better methods of disclosure, we can educate more people about our practices, then this is a valuable process.

That the study was carried out on a single platform, Apple's iOS, does raise an interesting point that the users of a particular operating system could be trained in what they understand to be private data, and therefore what should be protected, based on the dialogs which the operating system presents to users when access is requested by an application. For example, if an OS requests permission to access location or accelerometer data, but not the contents of media libraries or a fingerprint of all the applications installed on the system then this may lead users to believe that the former are more serious issues. What each system protects in this way is broadly similar, yet is changing as user expectation and consumer regulations evolve – although the concerns of users and the projected concerns of the system engineers who decide on these restrictions are not always in sync [8].

The work in this paper concerns research trials being conducted through mobile software. Many commercial applications might also be recording user information, of course, more commonly with the goal of improving their product or service rather than in answering research questions. We would suggest that these processes are broadly similar, but important differences occur in cases where researchers record additional information to answer a 'hidden' research agenda that is not related to the primary function of the app. At the core of this is user *expectation,* as we have previously identified [25] as being one of two main factors (along with personal identifiability of collected data) that we use to classify the risk associated with a trial, and the steps we recommend are taken to mitigate this risk. The study described in this paper fits into our classification as the recommended action for highest risk trials – interrupting users to show them collected data.

DELAYED T&CS FOR BETTER INFORMED CONSENT

Based on our explorations in this paper, we propose broadening our visual presentations of recorded information from merely location to all data logged during a trial, and using this as an alternative to first-launch T&Cs. In this trial, we showed T&Cs on first use, and uploaded data as soon as consent was agreed. In future, we propose that data initially remains on the device, until the later interruption takes place.

On installation, there would be a short period of data collection, with this information stored only on the local device. Once sufficient data had been collected, the user would be interrupted with a notification disclosing the fact that the application logs data on their usage. Visual presentations of each type of data would be displayed in the interruption, in the manner of the location data in this study. The nature of this visualisation would be determined by not only the type of data being collected but also the analysis to which it would be subjected by the researchers – showing the map of user location may not be sufficient if the intention was to analyse for colocation; a sample of colocation (possibly with synthetically generated data) would have to be provided. Tools would then be made available to the users to control which types of logging they wish to allow. If the user were willing, all data recorded so far would be uploaded to our servers. If the user declined, the data recorded so far would be deleted without ever having left the local device.

Our hope would be that these visual presentations would help users differentiate more between different types of data capture, and dissuade those with concerns from simply shutting off all forms of logging. It may be the case that users still find it easier to disable all forms of logging if they have concerns, rather than to go to the effort of making separate decisions for each type. We also intend to experiment with in-game rewards for each type of data, to explore what effect this has on participation rates.

We acknowledge that a limitation of our study is that all the forms of analysis we have undertaken were built into a game. It is possible that ‘gamers’ might have different views on privacy than the general population. We intend to run similar experiments in other forms of application. We also intend to run parallel trials with groups of local users, in order to conduct a richer qualitative analysis of the issues surrounding informed consent.

CONCLUSIONS

The HCI community has yet to reach a consensus on how best to gain informed consent from participants recruited via ‘app stores’. Our results indicate that it is incumbent on researchers to move beyond the current common practice of T&Cs and go to greater lengths to ensure that users understand and are willing to participate in their research.

We intend to continue to study larger groups of users, to look beyond location to other aspects of recorded data, and to explore at what stage of use it is most appropriate to interrupt users with presentations of recorded data. Our analysis of reported concerns, logging control alterations and text comments all support findings from the literature that the majority of users are relaxed about being recorded and allow their activity to be logged even when greater opportunities are provided to stop. However, a certain minority of users are unhappy with the process, reporting serious concerns and requesting logging to cease. It is perhaps this latter group of users that researchers should be going further to protect, by making more efforts to provide them with greater levels of understanding of the process, and therefore greater autonomy in participation.

Recent studies suggest that users are perhaps becoming more relaxed about data logging. We should never overestimate this movement in opinion, as the potential danger to the reputation of researchers or to the field in general outweighs the extra effort required to take a slightly more cautious approach. Additional measures beyond standard ‘read now or miss it’ T&Cs, providing simpler opt-out mechanisms and in particular showing the consequences of logging with real world examples might be helpful in ensuring that users’ consent is truly informed.

ACKNOWLEDGEMENTS

We thank Owain Brown and the other members of SUMgroup at the University of Glasgow, and acknowledge UK EPSRC funding (EP/F035586/1). An earlier version of this work appeared as a CHI 2011 Work in Progress [27].

REFERENCES

1. Böhme, R and Köpsell, S. “Trained to accept?: A field experiment on consent dialogs” *Proc. CHI’10*, 2403-2406 (2010).
2. Böhmer, M., Hecht, B., Schoning, J., Kruger, A., and Bauer, G. “Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage.” *Proc. MobileHCI '11,* 47–56 (2011).
3. Business Week/Harris Poll: “A Growing Threat” http://www.businessweek.com/2000/00_12/b3673010.htm (2000).
4. COM(2012).10 “Proposal For A Directive Of The European Parliament On The Protection Of Individuals With Regard To The Processing Of Personal Data … And The Free Movement Of Such Data.” *European Union,* (2012).
5. Cramer, H., Rost, M., and Bentley, F. “An Introduction to research in the Large.” *IJMHCI, 3*(4), (2011).
6. Egelman, S. “My Profile is My Password, Verify Me!: The Privacy/Convenience Tradeoff of Facebook Connect.” *Proc. CHI '13*, 2369–2378 (2013).
7. Federal Trade Commission. “Mobile Privacy Disclosures: Building Trust Through Transparency.” Federal Trade Commission Staff Report (2013).
8. Felt, A. P., Egelman, S., and Wagner, D. "I've got 99 problems, but vibration ain't one: a survey of

smartphone users' concerns." *Proc. SPSM'12,* 2012.

9. Friedman, B., Lin, P., and Miller, J. K. (2005). “Informed consent by design.” In *Designing secure systems that people can use*, L. Cranor and S. Garfinkel, Eds. O’Reilly, 495–521 (2005).
10. Good, N. S. et al. “Stopping spyware at the gate: a user study of privacy, notice and spyware.” *Proc. SOUPS*’05, 43 – 52 (2005).
11. Good, N. S., Grossklags, J., Mulligan, D. and Konstan, J., “Noticing notice: a large-scale experiment on the timing of software license agreements.” *Proc. CHI’07*, 607-616 (2007).
12. Grossklags, J. and Good, N. “Empirical studies on software notices to inform policy makers and usability designers.” *Proc. FC'07/USEC'07*, 341-355 (2007).
13. Henze, N., Pielot, M., Poppinga, B., Schinke, T., and Boll, S. “My App is an Experiment: Experience from User Studies in Mobile App Stores.” *IJMHCI, 3*(4), 71-91 (2011).
14. Henze, N., Rukzio, E., and Boll, S. “100,000,000 taps: analysis and improvement of touch performance in the large.” *Proc.Mobile HCI '11,* (2011).
15. Hudson, James M. and Amy Bruckman. "'Go Away': Participant Objections to Being Studied and the Ethics of Chatroom Research." *The Information Society* 20(2), 127-139 (2004).
16. Kay, M. and Terry, M. “Textured agreements: re-envisioning electronic consent.” *Proc. SOUPS’10* (2010).
17. Khovanskaya, V., Baumer, E. P. S., Cosley, D., Voida, S., and Gay, G. "Everybody Knows What You're Doing": A Critical Design Approach to Personal Informatics. *Proc. CHI '13*, 3403–3412 (2013).
18. Krumm, J. “A survey of computational location privacy” *Personal Ubiquitous Computing,* 13(6) 391-399 (2009)
19. Lee, H., Park, H., and Kim, J. “Why do people share their context information on Social Network Services?” *International Journal of Human-Computer Studies, 71*(9), 862 - 877 (2013).
20. Luger, E. and Rodden, T. “An Informed View on Consent for UbiComp.” *Proc. UbiComp '13*, 529–538 (2013).
21. Luger, E., Moran, S., and Rodden, T. “Consent for all: revealing the hidden complexity of terms and conditions.” *Proc. CHI '13*, 2687–2696 (2013).
22. Malhotra, N. K., Kim, S. S., and Agarwal, J. “Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model.” *Information Systems Research, 15*(4), 336-355 (2004).
23. McDonald, A. and Cranor, L. "Cost of reading privacy policies, the." *ISJLP* 4 (2008): 543.
24. McMillan, D., Morrison, A., and Chalmers, M. "A Comparison of Distribution Channels for Large-Scale Deployments of iOS Applications." *IJMHCI* 3(4) (2011): 1-17.
25. McMillan, D., Morrison, A., and Chalmers, M. “Categorised ethical guidelines for large scale mobile HCI.” *Proc. CHI '13*, 1853–1862 (2013).
26. McMillan, D., Morrison, A., Brown, O., and Hall, M. “Further into the wild: Running worldwide trials of mobile systems.” *Proc. Pervasive '10,* 210–227 (2010).
27. Morrison, A., Brown, O., McMillan, D., and Chalmers, M. “Informed consent and users' attitudes to logging in large scale trials.” *Proc. CHI EA '11*, 1501–1506 (2011).
28. Morrison, A., McMillan, D., Reeves, S., Sherwood, S., and Chalmers, M. “A Hybrid Mass Participation Approach to Mobile Software Trials.” *Proc. CHI’12,* 1311–1320 (2012).
29. Morrison, A., Reeves, S., McMillan, D. & Chalmers, M. “Experiences of mass participation in Ubicomp research.” Research in the Large Workshop, Ubicomp (2010).
30. Patrick, A. and Kenny, S. “From privacy legislation to interface design: Implementing information privacy in human-computer interactions.” *Privacy Enhancing Technologies,* 107–124 (2003).
31. Pielot, M. Henze, N. and Boll, S., “Experiments in App Stores - How to Ask Users for their Consent?,” Ethics, Logs and Videotape workshop, CHI (2011).
32. Strater, K., and Lipford, H. R. “Strategies and struggles with privacy in an online social networking community.*” Proc. BCS-HCI ’08*, 111–119 (2008).
33. The Guardian, “Path fined $800,000 by FTC over iOS privacy breach” http://www.guardian.co.uk/technology/2013/feb/01/path-fined-ftc (2013).
34. The Wall Street Journal, “The 0.00002% Privacy Solution” http://online.wsj.com/article/SB10001424052748704474804576222732361366712.html (2011).
35. Vihavainen, S., Karhu, K., and Botero, A. “Connecting Stakeholders Through Context Logging.” *Proc. MobileHCI '13*, 504–509 (2013).