Categorised Ethical Guidelines for Large Scale Mobile HCI

Donald McMillan
Mobile Life Centre @ SICS
SE-164, Kista, Sweden
don@mobilelifecentre.org

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

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

ABSTRACT
The recent rise in large scale trials of mobile software using 'app stores' has moved current researcher practice beyond available ethical guidelines. By surveying this recent and growing body of literature, as well as established professional principles adopted in psychology, we propose a set of ethical guidelines for large scale HCI user trials. These guidelines come in two parts: a set of general principles and a framework into which individual app store-based trials can be assessed and ethical concerns exposed. We categorise existing literature using our scheme, and explain how researchers could use our framework to classify their future user trials to determine ethical responsibility, and the steps required to meet these obligations.

Author Keywords
Ethics; large-scale trials; mass participation; app stores.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/methodology

INTRODUCTION
Large scale trials of mobile HCI systems using 'app stores' are becoming increasingly popular [9], yet it has been noted that such trials raise non-trivial ethical challenges [5]. Issues such as competing sets of guidelines, differences in international laws and research cultures, and lack of community consensus can leave researchers unsure as to how to run a study which meets their ethical obligations.

Perhaps the best-known guidelines specific to mobile and ubiquitous computing are those in Greenfield's Everyware book [18]. High-level guidelines such as 'do no harm' and 'default to harmlessness' were discussed, and are still generally applicable, but have yet to be contextualised to suit new ubicomp research practices. Emerging user practices, such as the widespread use of web sites such as YouTube and Facebook, and the near-ubiquity of cameras on phones also make more established guidelines, e.g. in MacKay's CHI '95 Ethics, Lies and Videotape paper [23] seem rather outdated. People are increasingly accustomed to the dissolution of traditional social barriers of privacy, driven by the poor privacy controls commonly provided by online social networking sites [36]. However, this shift in user attitude cannot be expected to be consistent across, or even within, cultures and demographics. The possible harm to the reputation of the researcher, or research as a whole, by overestimating this movement in opinion outweighs the benefits of taking a too relaxed attitude to the ethical diligence required by researchers.

In this paper, we look at existing sets of ethical principles for human trials, making a case for how they could be interpreted to encompass large scale mobile software trials. We then identify two key ethical issues for mobile HCI: anonymisation of participants and user understanding/expectation of data logging. We create a four-category classification of ethical requirements based on these factors. This is illustrated by surveying recent literature from this field and assigning past studies to one of our created categories. Finally, we take a two-stage approach to proposing a set of ethical guidelines for mobile HCI. Firstly, we provide a set of general principles based on our interpretations of the existing guidelines and secondly, we provide guidelines for each of the quadrants identified in our framework.

EXISTING ETHICAL GUIDELINES
Increasingly, HCI research crosses institutional, professional and national boundaries, further complicating the application of appropriate ethics protocols and review processes. For these reasons, researchers' development of detailed and specific regulations on the handling of ethics issues in HCI research, with the aim of covering all eventualities, is seen by many ethicists as an ultimately flawed direction [14]. As soon as one new set of regulations is finalised, a new method or topic of research is likely to emerge that is not covered. The existence of lengthy, detailed and prescriptive professional or institutional regulations raises the risk of researchers following the letter, but not the spirit, of the regulations and may in consequence lead to research being carried out that is ethically flawed.

HCI is by no means the only field of research which uses human trials as a method of evaluating hypotheses and exploring ideas. Notably, the fields of Psychology and Medicine have well established guidelines compiled and upheld by professional bodies. While it may be argued that the potential harm of an ill-run medical or psychological trial is much greater than that posed by a piece of mobile or online research, significant concerns are raised by the increasing value of personal data, the volume of such data it is possible to collect and the difficulties surrounding anonymisation.

CHI 2013, April 27–May 2, 2013, Paris, France.

BPS	APA
Respect for the Autonomy and Dignity of Persons.	Respect for People's Rights and Dignity.
Maximising Benefit and Minimising Harm.	Concern for Others' Welfare.
Social Responsibility.	Social Responsibility.
Scientific Value	Professional and Scientific Responsibility.
	Integrity.
	Competence.

Table 1: The Principles of Ethical Research from the BPS and APA.

Both the British Psychological Society (BPS) [41] and the American Psychological Association (APA) [35] provide principles-based guidelines for researchers conducting human trials. In using guidelines designed for psychologists as opposed to medical trials as a starting point, we believe that the resultant guidelines for mobile HCI will provide greater levels of ethical surety without unnecessarily restricting researchers in the field.

BPS and APA Guidelines: Interpretation for Mobile HCI

The BPS gives four guiding principles specifically for researchers using human subjects. The APA builds its ethical guidance for research from the six general principles given to cover all aspects of professional conduct. The titles of these are given in Table 1, ordered to show their similarity. While items in one column do not necessarily map directly to their counterparts in the other, there is a marked overlap.

For each of these guidelines, we will suggest how they can be interpreted for app store-based trials, and highlight the challenges they raise. Much of this discussion covers previously identified ethical issues in large scale mobile HCI trials [5], such as researchers not meeting participants recruited through an app store, the difficulty in accurately assessing age of participants, information provided by researchers to users not being read, and the inability to debrief following the trial.

1. Autonomy, Dignity and Self-Determination

The principles in the first row of Table 1 are described by the BPS and the APA as dealing with the 'dignity and worth of all' [41, 35]. Beyond that, they both discuss the fundamental rights that any participant has to self-determination and autonomy and personal liberty. In terms of HCI, it is therefore essential that researchers ensure the autonomy of the participants by verifying that they understand the consequences of participation, and that participation is optional. In order to ensure this, the *capacity* and *understanding* of participants must be verified, both of which pose a problem in the mobile space. Similarly, each participant's liberty should be protected, so researchers should safeguard the right to withdraw from a trial at any time and provide the knowledge and tools to do so.

Understanding

An important part of abiding by ethical standards in the running of a human trial is the nature of the agreement: the mutual understanding that a researcher is able to make with the users. In the context of a traditional trial, researchers can discuss the experiment, determine the participant's level of understanding and dynamically adjust the amount of information they present verbally to ensure that the most important points are clearly understood. If they are unable to bring the participant to a suitable level of understanding of the consequences of participating in the trial, they have a number of options available to them, from letting the participant complete the trial but deleting the data, to halting the process altogether and finding an alternative subject.

However, if a participant has downloaded trial software from an app store, researchers are not physically present to explain the trial. Therefore, a common procedure has emerged [24, 19] of presenting a briefing page of Terms and Conditions (T&Cs), and asking for confirmation of understanding and acceptance before allowing use. Such briefings at a distance over the Internet suffer from the removal of the subtle clues and cues that give the researcher extra information on the participant's level of understanding and the chance to reiterate and reword as needed. Further, the percentage of people who read T&C pages on installation of desktop software was reported by FAST [4] as being only 28%, and another study reported that only 20% of users who understood that the End User Licence Agreement (EULA) was a contract had any idea what it contained [17].

It is therefore doubtful whether this use of checkboxes and T&Cs pages can be said to gain truly informed consent for research applications, and so the extent to which the researcher can ethically collect data and publish the results of its analysis is equally questionable.

User understanding is further reinforced in traditional human trials during the debriefing stage. This is an important part of ethical practice, as it ensures that the understanding negotiated before the trial has stood up to the reality of participation. It offers participants the opportunity to ask for clarification on aspects of the trial or data collected that they were unable to fully grasp before participation. It also allows the researcher to gauge the effect that participation has, and spot any areas in the pre-trial briefing that need to be clarified, and is particularly important in any trial involving deception of the users. Unfortunately, as researchers of large scale trials will likely not meet their participants, and may have no means of communication with them except through the app itself, detecting the end of participation is an incredibly difficult task. On the four most popular smartphone operating systems (iOS, Android, Blackberry and WindowsPhone), the developer is not given the opportunity to interact with the user when an application is removed from the device; there is no equivalent to a desktop application's custom uninstall program. Even if uninstalls could be detected, the user may have ceased participation long before he or she gets round to deleting the app from the device. This could also occur after the analysis or publication of results based on their data has occurred, rendering any debriefing in such an uninstall program ineffective.

Capacity

The challenges of obtaining ethically valid informed consent are further complicated by considering whether a potential participant is even legally capable of providing consent. Determining that a participant has the legal capacity to enter into the contract set out by the researcher, at least in the UK, means determining that they are "an adult of sound mind".

Of course, even when physically present, determining that a participant meets the requirements of a given jurisdiction or for a given university ethics board is by no means foolproof. For example, a participant may make a false statement regarding his or her age for a variety of reasons, the most obvious of these being any reward offered for participation. Typical online solutions include 'confirm your age' check boxes, requesting the user to enter a date of birth or providing the details of a valid credit card. However, replacing the researchers' judgements on physical appearance, personality and validity of identification with any of these systems leads to a further marked degradation in the confidence that can be put in any given participant in a remote trial being of age, and of what the age of majority is where the participant resides.

Note that entering into contracts with minors or those not of sound mind is not an illegal action on the part of the researcher, but a contract made with a minor is, barring a few special circumstances, binding only to the adult or corporation – the minor may not face sanctions for failing to uphold their responsibilities according to the contract [6].

Opting Out

Standard research practice dictates that users should be able to cease participation in an experiment at any time, and researchers would delete all data collected on request. However, this data might have moved beyond the control of the experimenters. In particular, information that has been used within an application or community, blog or forum posts, or information that has been combined into the products of other users, such as mash-ups, raise significant problems. Beyond the purely practical challenges in deleting this data, the apparent ethical commitment to purge all data from one participant could be seen to cause harm to another.

Additionally, user-generated content might have been copied, commented on and published by others, with or without researchers' knowledge, e.g. via in-application sharing features, on their own blogs, or on social networking sites. The level of responsibility researchers have for such self-published information, and the validity of collecting it for analysis must be examined.

2. Concern for Others' Welfare

The second of the principles in Table 1 of 'Maximising Benefit and Minimising Harm' & 'Concern for Others' Welfare' both deal with the theme of risk. It is the researchers' ethical responsibility, as a general rule, not to expose the user to any risks greater than they would encounter in their everyday lives. They should be aware of the real and perceived power differences in their relationship with the participants and be careful not to exploit their research subjects. In research which poses risks to the participants' psychological well-being, mental health, personal values, or dignity, these risks should be assessed to determine their probability and severity, and measures should be put in place to minimise the exposure of the participants and to recover should the worst-case scenario be realised.

In this regard the BPS points out that the responsibility of the researcher goes beyond direct harm that may be caused and that they must be 'alert to the possible consequences of unexpected as well as predicted outcomes of their work.'

Anonymisation and Re-identification

One of these unexpected consequences might be tracing anonymous data back to its creator. Although attempts may have been made by researchers to record and store data in a non-personally identifiable manner, there have been many cases of subsequent re-identification of notionally anonymised data. In 2000, Sweeney [37] showed that 87 percent of all Americans could be uniquely identified using only three pieces of information: their postal code, birthdate, and sex. In doing so she was able to take the 'anonymised' data released by the Massachusetts Group Insurance Commission on all their state employees and, when combined with the 'anonymised' voter rolls from the city of Cambridge, which were purchased for $20, identify the current state governor's health records including his diagnoses and prescriptions. Similarly, students at MIT cross-referenced the Chicago Homicide Database with the Social Security Death Index to re-identify the victims of homicides and whether the murder involved drugs, child abuse, gang violence, or domestic abuse as well as previous criminal history of the victim [29].

This work, and others like it, show that almost all information can be defined as 'personal' when combined with enough other relevant data. Additionally, new potential identification methods are constantly being developed, for example by collecting and analysing hand tremors [21]. With the continual advancement of identification and re-identification techniques, what information can, and can't, be said to be anonymous or insignificant continually changes.

3. Social Responsibility

The third row of Table 1 shows that both the BPS and the APA feel that Social Responsibility is central in deciding if a course of action is ethical. As a researcher, one must be mindful of and responsible to the societies in which one lives and works.

Societies have already created several laws to govern many of these practices. Researchers, by storing any personal data related to their participants, are already legally bound in many countries by legislation such as the EU directive on the Protection of Personal Information [10] which is soon to be augmented by the EU regulation on the Processing Of Personal Data And On The Free Movement Of Such Data [8]. As shown on Forrester's Privacy and Data Protection by Country Heatmap[1], the majority of the world's population is covered

[1] heatmap.forrestertools.com

by some form of data protection law and researchers must ensure compliance with the law where they live and work.

Sharing raw or anonymised data between researchers and institutions is standard practice in many fields – ensuring that results are reproducible and that the greatest amount of knowledge can be extracted from the effort expended in gathering such corpora. Yet, as outlined earlier, successful anonymity of data is increasingly difficult to achieve.

The recently passed EU directive on the movement of personal data [8] introduces a principle that could result in more ethical practice when data is shared between researchers. Here an entire chapter is devoted to the 'Rights of the Data Subject' obliging *data controllers*, those collecting and processing the data, to provide transparent, easily accessible and understandable information on what has been collected and to provide procedures and defined deadlines for requests for access, and deletion, of personal data. Of interest with regards to the sharing of data between researchers is the chapter that deals with the transfer, or onward transfer, of data outside of the EU or to an international company not governed by this directive. It states that the collector of the data is responsible for ensuring that it is not transferred to another party with less stringent security protocols.

4. Scientific Value, Integrity and Competence

The principle of Scientific Value put forward by the BPS incorporates many aspects of the principles of Professional and Scientific Responsibility, Integrity and Competence put forward by the APA. The scientific value of the research must be clear and appropriate (Professional and Scientific Responsibility), the research must be well designed and conducted in a way that ensures its quality (Competence) and integrity. For large scale mobile HCI systems, we would suggest that relevant issues here are protection of user data – on the device where it is collected, in transmission to researchers' servers and in storage in the researchers' database. The APA principles add, in regards to research, that the practitioner has a responsibility to intervene with colleagues to prevent or avoid unethical conduct.

Internet Mediated Research

Further to their general guidelines, the BPS provide a supplementary publication dealing with the specifics of conducting research online [40]. These guidelines for Internet Mediated Research (IMR) present a set of recommendations, but this work predates app stores and the points raised are, in most cases, specific to the research cases that were examined – mainly observation of online forums and running online surveys. Therefore, many of the recommendations are not directly applicable to the methods of research discussed here. However, the IMR publication does identify two key issues where new ethical problems are faced when using remote participants – the identifiability of participants and participants' levels of understanding that they are part of the research. In the following section, we use an adapted version of these identified key pair of principles as the basis for our ethical framework for large scale mobile HCI trials.

ETHICAL FRAMEWORK FOR LARGE SCALE MOBILE HCI

In this section, we describe a framework that can be used to make a proportional classification of mobile HCI studies, and identify the ethical responsibilities of each particular user trial. Related to the key issues highlighted in the general IMR guidelines, we identify 2 key dimensions as being of particular importance to mobile HCI studies, and create a 2 dimensional representation into which trials can be classified. We argue that this schema is of benefit as it allows a proportional view of user trials and researcher practice, where we are conscious that broad 'one size fits all' guidelines can fail to take into account the subtle challenges of real world studies, and that the steps taken to meet ethical obligations should reflect the explicit challenges and risks involved in the particular user trial being conducted.

To illustrate our framework, we categorise several past studies from the literature into our schema. Thereafter, in the Guidelines section, we set out recommendations to specifically address the challenges of each of our identified categories.

Key Dimensions: User Expectation and Identifiability

As previously explained, it is a fundamental ethical responsibility of all research trials not to expose users to greater risk than they would ordinarily encounter. We note that even before large scale trials and app stores, mobile HCI research applications and trials could have involved risk. Examples might include encouraging participants to carelessly cross busy streets or to lower security settings on hardware that would increase their chances of exposure to malware. However, in this work, we are concerned only with those risks specific to running an app store-based trial, where researchers must take more care as a result of operating at a large scale than they would have when interacting with a local group.

We investigated a number of different risk factors that could be increased by the combination of this type of remote trial and the system or trial design. We considered classifying based on issues such as locatability of individuals, access to personal data, and the researchers' intended forms of analysis (e.g. analysing in aggregate vs. studying individuals in detail) [28]. However, based on trends in the literature and our own experience of running several mass participation trials, we uncovered two key issues with which we felt able to sensibly classify published research systems.

The first of our identified key dimensions is user expectation. In a more traditional trial process, the expectations of the user with regard to the behaviour of an application and of the researchers would be negotiated face-to-face, to ensure understanding. However, as mentioned before, without this there is a greater risk of breaching the autonomy of the subject and a risk, if such a breach is uncovered by the subject, of causing damage to the reputation of the research field as a whole.

If the T&Cs screens presented to users when an application is installed or launched were always read and understood, researchers would gather truly informed consent, as in a more standard face-to-face trial, and these additional risk factors would be mitigated. However, it has been shown that users can ostensibly agree, but not actually read these screens. In

such cases the user is agreeing to the behaviour of an application which matches his or her mental model of how such an application *should* behave. Therefore, the degree to which the software matches a user's *expectations* of the kind of data being accessed, recorded and shared is of key importance.

For example, it can be argued that when a participant downloads a maps application that relies upon user location for its central functionality, she would expect her location to be exposed to the application, and by proxy the developer of the service being used. This expectation would not be there when downloading a calculator application. In our previous research in this area, we experimented with recording location in a game that did not use location functionality, and later interrupting users to inform them of this [27]. Even though we had disclosed our logging policies in T&Cs that had to be accepted before use of the game could begin, we noted strong reactions from users who were very surprised that we were capturing this information that seemed extraneous to the application. Comments included "U should not hav the right to do this!!!" and "it's invading in peoples privacy :(".

There is a precedent for seeing this as an important factor. One of the 6 principles Friedman et al [12] present to guide the design of consent interfaces concerns "minimal distraction", where they advise against asking for specific consent from users too often, as this would lead to 'interruption fatigue', and instead to rely on the implicit consent given by virtue of entering into a situation where such activities are broadly known to occur. Friedman et al recommend saving interruptions for cases where user expectation would be broken, highlighting the significance of this issue.

The second key dimension identified was the level of anonymity afforded to the user by the data collection process. Uniquely identifying a user greatly increases the risk of embarrassment or reputational harm, particularly if the user believes she was acting or posting comments anonymously. Iachello and Hong identify the management of personally identifiable information as a key factor both in personal privacy and in the perceived risks which influence a user's digital privacy preferences [20]. These risks are particularly high in a large scale trial because of the volume of data collected, the potential visibility of the collection and the number of people who could be harmed by a breach in security or the re-identification of released but thought to be anonymous data.

Figure 1 illustrates these two key dimensions. The *y*-axis can be seen as a measure of the disconnect between what the application does with regards to the users' data and the mental model of the application the researcher can reasonably expect the users to have. The *x*-axis is the identifiability of the data being collected. The combined risk of these two factors therefore increases as the data logged in a trial becomes less in keeping with users' expectations and as participants become more identifiable by the data captured by the researchers, and we suggest that more stringent ethical directives are required as both of these factors increase.

To illustrate our framework, we explain each of the four quadrants of Figure 1 in turn, and classify a representative cross

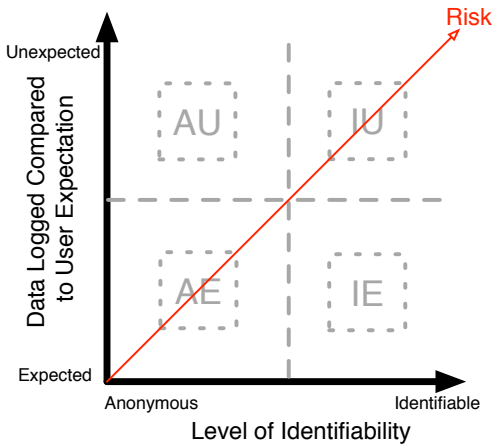

Figure 1. Two key dimensions of ethical responsibilities: users' expectations of data usage and the level of anonymity afforded by the data logging, storage and transmission processes of the researchers. Four quadrants are identified, e.g. AE = anonymous, expected

section of published research, to help explain each of the identified quadrants and to show how existing works would fit into our categorisation. Following this, we provide a number of ethical guidelines that we would recommend for future trials in each section.

Readers should note that the categorisation below is taken solely from information explicitly reported in publications. It is impossible to determine from these publications the true extent of the logging being performed, as some forms of exploratory research might involve recording as much data as possible when research questions have not been clearly defined, or as a safeguard against failing to capture something important.

Anonymous data, Expected collection (AE)

For a trial to be categorised into this quadrant researchers would only be logging information which the user would reasonably expect to be made public during the use of an app, such as data that is obviously vital to the software's functionality. Also, that data would pose only a small risk of identifying the user individually.

Examples of releases in this category are those for which only the number of downloads is reported, such as the 5 localised botanical identification apps by Riccamboni [33] covering different locations and price points. Walk'n'Play [3], an app that measured the amount of calories expended by each user, reported only the number of users and the aggregate length of engagement with the application.

ZooEscape and Packer[25] reported download numbers as well as the self-reported demographic makeup of the users, information that, while providing a slightly higher risk of identification than numbers of downloads, is still low enough to fall into this initial quadrant.

Other members of this quadrant would be applications released to collect a specific form of data which is an integral part of the application, and yet not easily used to identify an

individual. Examples include Text Text Revolution [34], a game that collects typing behaviour and error rate, Hit-It [19], a game where users tap moving targets to collect accuracy statistics for targets of different size and position, and OverCharged [11] that gave users information on and collected their device's battery level.

Anonymous data, Unexpected collection (AU)

Applications that fall into this section also gather data which is unlikely to expose the users' identity. However, in these cases the information is, to a greater extent than those in quadrant AE, getting further removed from what a user would reasonably expect to be recorded.

Oliver [30] released a logger to investigate how users interact with and consume energy on their portable devices. Where this differs from Ferrera's work with OverCharged, assigned to quadrant AE, is that this logger was 'hidden' within another, unrelated application. In much the same way an application for a leading Swiss insurance firm was used to deploy logging to examine the location accuracy of iOS devices [39].

Nokia Research Centre's Friend View [7] reported statistical analysis of social network patterns. While, in general, the access of a user's social graph would be seen as collecting data which is highly identifiable and would present more risk, in this case the application anonymised the data before the researchers analysed it.

Identifiable data, Expected collection (IE)

Applications which fall into quadrant IE are those for which the user would expect the data collected to be accessible by the developer, such as data that is necessary for the operation of central functionality in the application itself, yet holds a greater risk of personally identifying a single user.

Perhaps the most obvious piece of collected data that would cause an application to be categorised into this quadrant is user location. PocketNavigator [32], exemplifying most navigation applications, falls into this category. Appazaar [2], AppAware [15] and Appjoy [42] detect users' app usage and location to recommend apps. Interestingly, the recommendation algorithm used in Appjoy would result in it being categorised in quadrant AE as the location data is not used; however it is collected for possible future analysis.

AppAware also collects users' social network accounts as does Cenceme [26], an application that uses context sensing to update social networking sites with users' activities. A user's social network account is a highly identifiable piece of data. Applications which collect information of such fidelity, such as real names, telephone numbers and to a lesser extent email addresses, are also categorised in this quadrant.

Identifiable data, Unexpected collection (IU)

Quadrant IU identifies those applications which collect highly identifiable data beyond the reasonable expectations of the users. This is the quadrant in which the applications pose the highest risk.

Twiphone [16] is an application that shares a user's call logs and SMS messages on Twitter. The data shared is highly personal and highly identifiable. While this behaviour may be expected by the user who installs the application, anyone *else* who sends an SMS to a Twiphone user may have that message unexpectedly and automatically broadcast on Twitter.

Hungry Yoshi [24] was a location based game which collected the location of its users, yet the location collected via GPS was not used as part of the game. This application also collected social networking account details.

GUIDELINES FOR LARGE SCALE MOBILE HCI

We now propose a set of guidelines, split into two parts. Firstly, having interpreted existing BPS and APA principles and noted the challenges raised, we present a set of general guidelines to address these points, which we suggest should be applied to all research undertaken in this area. The second set of guidelines recognises that one size doesn't fit all and uses the two-dimensional framework introduced in the previous section to consider the behaviour of both the application and the researcher, providing more tailored guidance on risk and responsibility.

General Responsibilities

This section describes the principles-based responsibilities of researchers, where we reflect on the points outlined in our interpretation of existing guidelines for psychology, and give our recommendations on the practice HCI researchers should adopt to satisfy these principles.

Regarding Autonomy

Many of the problems surrounding autonomy concern informed consent, and users' capacity to provide it. While this remains a difficult area, we suggest that researchers must take care not to intentionally or unintentionally target vulnerable groups when advertising their trial. The recruitment of participants for remote mobile trials can take many forms, from physical fliers to demographically-targeted online advertisements, but the most basic form is the icon, description, keywords and chosen categories used in the online repository. Where the store allows, the researcher should restrict the application to those over the age their institution has deemed acceptable for participation and, if the trial would not be adversely affected, it would be advisable to raise this limit to an internationally recognised age of majority such as 16, 18 or 21. Icons, screenshots and description language can also be created in such a way that researchers can appeal to their target audience and not unintentionally target children.

The information needed to provide consent should be given in T&Cs, available within the application itself as part of the 'help' or 'info' screens, and should also be part of the online description to give users the best possible chance of realising that the software they are about to install is part of a research trial. Due to the global nature of this distribution method we recommend that this information be presented using simple, easy to understand language and that it be available in each localisation that the application itself is. We acknowledge, however, that users will still often fail to

read T&Cs; we outline further mitigating procedures in the section on categorised guidelines.

In keeping with the principle of rewarding research participants for their cooperation, researchers can restrict functionality to those who have agreed to participate. For example, they could choose to restrict location-based services to those willing to share their location or restrict social networking functionality to those willing to share demographic information.

If this approach is taken then the socio-cultural pressure applied should be monitored and, in the event that the pressure of a peer group using a service or application is suspected of impacting a participant's agency, access should be made available without collecting and analysing the data.

We would suggest that issues surrounding opting out of an experiment or dealing with requests to delete data can be solved through careful experimental and software design. Where the trial and the application have been correctly presented to users, there should be little ambiguity as to what content the participant is explicitly sharing, and with whom. In situations where it can be reasonably expected that the sharing would be subject to few limits, such as uploading a video to YouTube or creating a custom blog theme that others have used, then the researcher should not feel under obligation to take action that would harm derivative content if a request for withdrawal is received. Such derivations can also be used in research. Where the participant has an expectation of privacy within the space that the content is shared, be that a Facebook post assumed to be limited to friends or a post in an online forum expected to be read by community members only, then the researcher must respect that expectation and work within its limits.

Regarding Risk

As previously outlined, any collection of identifiable or re-identifiable data could cause potential harm and therefore must be stored and transferred securely. While it cannot be expected that each research group, or individual researcher, be at the forefront of data security and encryption research, they have an ethical duty to keep abreast of the current industry standards in this area. They must also take steps to mitigate risks that are identified in the planning and in the course of their research, including the risks resulting from them not being security professionals, by instigating suitable procedures with regards to their handling of the data.

All trial data on the participant's device should be encrypted and deleted after being successfully transferred back to the researcher or when the participant withdraws from the trial. All cached data should be deleted when uninstalling the software or withdrawing consent from within the application. This mitigates the risk of personal historical data being made available to persons of ill intent with physical access to the device, or other rogue software processes.

The externally visible server which receives the uploaded data from the client software provides a possible point of failure which could affect a large number, if not all, of the trial participants. Beyond ensuring that the operating system and web-server software is regularly updated, ensuring that passwords are of an appropriate complexity, and ensuring that access permissions are correctly set, the responsible researcher should recognise that the risk of a security breach is still there and minimise the damage it would cause. One way to do this would be to avoid keeping the database of historical log data on such an externally visible server – by regularly moving the incoming user data to another fully fire-walled or offline machine, the potential amount of data compromised by a breach is greatly reduced.

As explained above, the sharing of data is often a desirable practice, but the recently passed EU directive on the movement of personal data [8] states that the collector of the data is responsible for ensuring that it is not transferred to another party with less stringent security protocols. Adapting this as a guideline for researchers would mean that the risk of data exposure through a security breach should be *at least* as low in the second institution as in the first, and that the original researcher is responsible for ensuring that this is the case before transferring the data.

Before any data is transferred outwith the control of the data collector, it should be subjected to a Privacy-Preserving Data Publishing technique – a survey of the state of the art of such methods can be read in [13] – which transforms the data by replacing any explicit or quasi-identifiers in the original with new identifiers that hide some detailed information so that several records become indistinguishable in this respect. This is, necessarily, reducing the fidelity of the data transferred so the researcher must take into account their trust in both the integrity and the security practices of the receiving researcher when deciding to what extent to employ these techniques. They should also discuss with the researcher requesting data its expected use and remove any fields not directly relevant.

Categorised Ethical Guidelines

In addition to these broad general principles, we also use our ethical framework for mobile HCI to allow researchers to determine the ethical responsibilities specific to their particular user trial. Having identified two key issues, and categorised the space accordingly, we now propose a set of guidelines for each quadrant. These rules would seek to supplement rather than replace the guidelines based on the BPS and APA ethical principles of the previous section. These guidelines are intended to be of increasing stringency, requiring further mitigating action on the part of the researcher as the trial is deemed to be classified further from the origin on either axis.

'Low risk' cases (Quadrant AE)

These would be applications classified in quadrant AE of figure 1, where an application only collects data that the end user would reasonably expect to be making available to the developer – such as data that is obviously vital to the software's functionality – and the collected data has a low chance of being used to identify the user individually. In such cases, the users' expectation of privacy is not being breached and the risk of identification is low, and so we suggest that the general guidelines suffice.

Dealing with identifiable data (IE)
Applications which collect identifiable data, which the user would expect to make visible to the developer during the routine use of the software, are encouraged to show the collected data within the app and offer clear controls to delete data.

Reflecting the higher risk associated with this type of data, users should be given the opportunity to review the data they have generated, and expunge any part of it. Such a procedure can be seen to take on the role of the debrief in a traditional human trial, giving users a greater understanding of what data has been collected, contextualised by their participation, and allowing them to withdraw.

It is important that this information is presented in a readable, accessible form. Taking as an example a location-based application, the researcher could show a clustered map of recorded locations or, using reverse geo-coding, a list of the most frequented street addresses. Simply showing a long list of GPS coordinates would be a less satisfactory solution.

Data Sanitisation procedures, an overview of which is provided by Bishop et al [1], can be employed to lower the identifiability of the data collected, lowering the risk and therefore the effort necessary to mitigate that risk. Data can be sanitised in one of two ways; perturbation and generalisation. In perturbation schemes, incorrect data values replace correct ones in such a way that the analysis of the perturbed data produces the same results as the original data. For example, if the domain is a set of numbers such as salaries, the perturbations must preserve the statistical moments of interest to the analyst. In generalisation schemes, the values are replaced by ranges that include the correct values. For example, replacing a birth date with the birth year replaces a precise value for a date with a range of values that includes that date.

Dealing with unexpected collection (AU)
In cases of applications that collect data the user would not reasonably expect to be collected, given the apparent functionality of the application, it is recommended that users be presented with an inline alert asking them to confirm consent that each of these unexpected data streams be shared with the researcher.

These 'Just-In-time Click-Through Agreements' (JITCTAs) [31] proposed by Patrick and Kenny in 2003 have begun to be built into the operating systems of many popular smartphones. For example, iOS has historically required specific user permission for an app to access location data, and as of the 2012 release of iOS 6 has extended this to include contacts and other personal information. Researchers extrapolating contextual information from sensors not protected by the operating system would be encouraged to display a JITCTA when the user performs an action that causes personal information to be collected by the software, or when the application firsts detects a type of personal information. The designer should be careful not to overload the user with requests as they start the application, in keeping with Friedman et al's principal of minimal distraction [12]. These JITCTAs should either be combined, or the collection of data that causes such JITCTAs to be presented to the user should be staggered.

Dealing with identifiable data & unexpected collection (IU)
In the IU quadrant are the applications which present the most challenging ethical dilemmas – those that collect identifiable data that the user would not reasonably expect to be collected by their interaction with the application given its apparent functionality. Since in these cases the data logging is both identifiable *and* unexpected, we recommend the use of both sets of procedures described above for dealing with identifiable data and unexpected collection.

Yet, because of the increased risk of their combination, we recommend that researchers go further in these cases. Therefore, we suggest that summaries of logged identifiable information are not only passively made available, but that participants are *actively interrupted* while using the application with the presentation of this data.

While this may be expected to have a detrimental effect on use and continued interactions, we have found [27] that displaying such representations of identifiable data may have negligible impact on these factors, and others [38] have seen an increase in use as a result of presenting such feedback.

Discussion
There are many cases where researchers could reduce the risk associated with their trial. For example, consider a study where the research question is determining the number of different locations in which an application is used. The simplest implementation of this might be to upload location data each time the user launches the application, then run a database query on the server to count locations for each user. This strategy entails capturing and transmitting a lot of potentially identifying information, and could see the trial classified in quadrant IU. However, an alternative approach is possible, where the recorded launch locations are stored locally on the device and the only information uploaded to the researchers is the number of different locations detected. With such an approach, the trial could move from quadrant IU to quadrant AU. By designing the application to include location-based functionality, it could be in quadrant AE.

A researcher, therefore, has options when planning a study. If it is necessary to have a record of all the raw information, we would recommend that the guidelines for quadrant IU are followed. Alternatively, a trial can be designed in such a way as to move closer to the origin of figure 1, and reduce the number of guidelines that would apply. We encourage researchers to think carefully about such issues, and avoid an attitude or culture where capturing, transmitting and storing as much information as possible becomes commonplace.

Neither of the dimensions of our framework are static measurements, and set rules cannot be ascribed for positioning a trial on either axis. We have explained that there are many subtle ways of potentially identifying or re-identifying an individual, which in themselves form an active area of research in Computer Science; what may today seem like anonymous data might tomorrow be personally identifiable. Similarly, users' expectations of the logging behaviour of applications is bound to change over time as technology and use evolves or due to changes in public perception of the value and ac-

cessibility of personal data caused by articles in the press or popular fiction. As noted by Iachello and Hong [20], expectation evolves with the use and adoption of technology, and is a reflection of the ongoing and organic "boundary definition process", by which people negotiate their disclosure of identity. During a long term trial of a novel technology this expectation will change, indeed actively designing this 'adoption pattern' is recommended. However, the proportional framework allows for this change, as the interpretation of risk on these dimensions is for the researcher to make at the time the trial is to be run and finessed if necessary as new members join the participant pool.

Further, we note that the diagram's sharp partitioning is used as a rhetorical tool for discussing the challenges and responsibilities that movement along these axes represents; the point at which the identifiability of the data being collected or the breach in perceived user expectations requires additional action is left to the researcher to determine.

At present we treat the population of users as if they have a uniform level of expectation, and defensively make a conservative estimate of that level. However, it may become possible to characterise the individual user in terms of activity and knowledge so as to apply the appropriate ethical guidelines or collect a different granularity of data accordingly. Work in this vein has been suggested by Lin et al [22], who propose crowdsourcing to capture people's mental models of an app's privacy-related behaviours.

We would advise that caution be applied in making these judgements. Not all within a user community will have the same expectations, and those expectations may vary depending on the context of the user. Similarly, what is potentially identifiable may vary depending upon the users' actions. People with regular patterns of commuter travel will have their homes and workplaces more easily identified than a travelling salesman.

CONCLUSIONS AND FUTURE WORK

The time is ripe for reconsideration of established research norms and practices, and researchers' understanding of public practices and sensitivities, so as to strike a new balance between invasiveness and utility. There are many ethical challenges being faced by researchers in many fields involving human trials as a result of the fast pace of technological advancement and incorporation into our everyday lives. These challenges come along with a number of exciting opportunities to use these new technologies to inform not only the design of the novel, but the understanding of the mundane. Understanding how we researchers can use this technology in ways which allow us to answer new and old questions with new levels of validity without harming the moral integrity of the community is a goal that should be pursued.

We have provided a set of general guidelines combined with a proportional framework to allow researchers to judge the ethical obligations and challenges posed by each individual trial. Practical examples of methods to meet these obligations are given and the tools that have been explored by the HCI community to achieve some of these goals have been highlighted. In general, we suggest that when deploying an application, researchers collect as little identifiable data as they can while still meeting their research goals, and do so in a manner transparent to the end user. Terms and Conditions pages have been shown to be ineffectual, so we argue that lowering risk by taking extra precautions to protect anonymity is a better approach than logging all available information, believing yourself to be 'covered' by ostensible user agreement.

Researchers should make every effort to inform users that their app is research software, so that informed consent can be obtained before data collection begins. Ultimately, it would be ideal for app stores to provide a 'research apps' category. End users would know that downloading any software from this section of the store would enrol them as participants in an experiment, and they could browse this section for the particular studies they were interested in contributing to.

There remains a question to be answered as to the level of control to give the user over the data being logged, and the cost to the user of withdrawing some or all of their data. We should consider cases where an application was run in error or before the user had understood the logging was taking place and provide tools with which users should be able to redact certain locations from the record in order to protect their privacy, without being forced to withdraw from the trial altogether. Yet this reduces the value of the data that the user is notionally 'paying' with to get access to the researchers' application. If the ability to delete or falsify some of their data is given to the user, should this be limited to a certain percentage before the only option available becomes to delete all data and stop using the app?

We are presently working towards the release and trial of applications which allow users to view, alter and report on the data they are providing, as well as the analysis and conclusions that we, as researchers, draw from this raw data. We are building a framework to allow us, and potentially other researchers working on iOS, to add this functionality with the minimal effort, along with two applications which will put this into the hands of users. The plan is to perform A-B testing by presenting different visualisations and notification methods to end users and comparing the results of their self-reported comfort with the system after seeing this data.

In parallel with this development work, we plan to conduct a two-stage enquiry into the attitudes and practices of researchers, users and commercial developers in this space. The first stage consists of a set of tailored surveys presented to members of each of these groups and covering as wide a demographic and geographic spread as possible. It will include questions on current practice and ask participants to report their opinions on a small set of observed and hypothetical application behaviours. The results of the analysis of this survey data will be used to inform the running of a series of workshops with each of the participant populations in order to gain detailed understanding of the results of the survey.

We hope that the guidelines presented here will help researchers to identify and combat the ethical challenges arising as they move to take advantage of the new opportunities af-

forded by app stores and large scale trials. Of course, we still rely on the judgment of researchers in identifying and overcoming the specific ethical challenges they face. We hope that the discussion on the ethical responsibilities of those conducting human trials with remote participants will continue, and involve research practitioners from across HCI and beyond. We hope that these guidelines will provide a solid base for that task, and encourage discussion towards the creation of a community consensus on ethical practice in large scale mobile HCI.

ACKNOWLEDGMENTS

We thank the other members of SUMgroup at the University of Glasgow for their collaboration, and acknowledge UK EPSRC funding (EP/F035586/1).

REFERENCES

1. Bishop, M., Cummins, J., Peisert, S., Singh, A., Bhumiratana, B., Agarwal, D., Frincke, D., and Hogarth, M. Relationships and data sanitization: a study in scarlet. In *NSPW '10*, ACM (Sept. 2010).
2. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., and Bauer, G. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *MobileHCI '11*, ACM (2011), 47–56.
3. Buddharaju, P., Fujiki, Y., Pavlidis, I., and Akleman, E. A novel way to conduct human studies and do some good. In *CHI EA '10*, ACM (2010), 4699–4702.
4. Campbell-Burt, R. Federation Asks: Do you know what you're agreeing to? Federation Against Software Theft, Mar. 2006.
5. Chalmers, M., McMillan, D., Morrison, A., Cramer, H., Rost, M., and Mackay, W. Ethics, logs and videotape: ethics in large scale user trials and user generated content. In *CHI EA '10*, ACM (2011), 2421–2424.
6. Chen-Wishart, M. *Contract Law*, 3 ed. Oxford University Press, July 2012.
7. Chin, A. Finding Cohesive Subgroups and Relevant Members in the Nokia Friend View Mobile Social Network. In *CSE '09* (2009), 278–283.
8. 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* (Jan. 2012).
9. Cramer, H., Rost, M., and Bentley, F. An Introduction to research in the Large. *IJMHCI 3*, 4 (2011).
10. Directive, E. 95/46/EC-The Data Protection Directive. *Official Journal of the European Communities* (1995).
11. Ferreira, D., Dey, A., and Kostakos, V. Understanding human-smartphone concerns: a study of battery life. In *Pervasive '11*, Springer (2011), 19–33.
12. Friedman, B., Lin, P., and Miller, J. K. Informed consent by design. In *Designing secure systems that people can use*, L. Cranor and S. Garfinkel, Eds. O'Reilly, 2005, 495–521.
13. Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys 42*, 4 (2010), 1–53.
14. Gilman, S. C. Ethics codes and codes of conduct as tools for promoting an ethical and professional public service: comparative successes and lessons. *Prepared for the PREM, the World Bank* (2005).
15. Girardello, A. AppAware: which mobile applications are hot? In *MobileHCI '10* (2010), 431–434.
16. Girardello, A., and Michahelles, F. Twiphone: Sharing communication behavior on twitter. In *IoT for Citizen Workshop @ Pervasive '10* (2010).
17. Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D., and Konstan, J. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *SOUPS '05*, ACM (2005), 43–52.
18. Greenfield, A. *Everyware: The Dawning Age of Ubiquitous Computing*. Peachpit Press, Mar. 2006.
19. Henze, N., Rukzio, E., and Boll, S. 100,000,000 taps: analysis and improvement of touch performance in the large. In *MobileHCI '11*, ACM (Aug. 2011).
20. Iachello, G., and Hong, J. End-User Privacy in Human-Computer Interaction. *FNT in HCI 1*, 1 (2007), 1–137.
21. Liberty, M. G., Roller, C. D., Simpkins, D. S., and Gritton, C. W. K. Methods and devices for identifying users based on tremor. Patent No. US7236156, Filed May, 2005, Issued June, 2007.
22. Lin, J., Amini, S., Hong, J., Sadeh, N., Lindqvist, J., and Zhang, J. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. *Ubicomp '12* (2012).
23. Mackay, W. Ethics, lies and videotape. . . . In *CHI '95* (1995), 138–145.
24. McMillan, D., Morrison, A., Brown, O., and Hall, M. Further into the wild: Running worldwide trials of mobile systems. In *Pervasive '10* (2010), 210–227.
25. 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.
26. Miluzzo, E., Lane, N., Lu, H., and Campbell, A. Research in the app store era: Experiences from the cenceme app deployment on the iphone. In *LARGE '10* (2010).
27. Morrison, A., Brown, O., McMillan, D., and Chalmers, M. Informed consent and users' attitudes to logging in large scale trials. In *CHI EA '10*, ACM (2011), 1501–1506.
28. Morrison, A., McMillan, D., Reeves, S., Sherwood, S., and Chalmers, M. A Hybrid Mass Participation Approach to Mobile Software Trials. *CHI '12* (Jan. 2012), 1311–1320.
29. Ochoa, S., Rasmussen, J., Robson, C., and Salib, M. Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study. *Unpublished Student Paper* (2001).
30. Oliver, E., and Keshav, S. An empirical approach to smartphone energy level prediction. In *UBICOMP '11*, ACM (2011), 345–354.
31. Patrick, A., and Kenny, S. From privacy legislation to interface design: Implementing information privacy in human-computer interactions. *Privacy Enhancing Technologies* (2003), 107–124.
32. Poppinga, B., Pielot, M., Henze, N., and Boll, S. Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application. In *WOMUE '11* (2010), 41–43.
33. Riccamboni, R., Mereu, A., and Boscarol, C. Keys to Nature: A test on the iPhone market. In *Tools for Identifying Biodiversity: Progress and Problems*, EUT Edizioni Università di Trieste (2010), 445–450.
34. Rudchenko, D., Paek, T., and Badger, E. Text text revolution: a game that improves text entry on mobile touchscreen keyboards. *Pervasive '11* (2011), 206–213.
35. Sales, B., and Folkman, S. *Ethics In Research With Human Participants*. American Psychological Association, 2000.
36. Strater, K., and Lipford, H. R. Strategies and struggles with privacy in an online social networking community. In *BCS-HCI '08*, British Computer Society (Sept. 2008), 111–119.
37. Sweeney, L. Simple Demographics Often Identify People Uniquely. *Carnegie Mellon University, Data Privacy Working Paper 3* (2000).
38. Tsai, J. Y., Kelley, P., Drielsma, P., Cranor, L., Hong, J., and Sadeh, N. Who's Viewed You? The Impact of Feedback in a Mobile Location Sharing System. In *CHI '09* (2009).
39. von Watzdorf, S., and Michahelles, F. Accuracy of positioning data on smartphones. In *LocWeb '10*, ACM (Nov. 2010).
40. Working Party on Conducting Research on the internet. Guidelines for ethical practice in psychological research online . The British Psychological Society, July 2007.
41. Working Party on Ethical Guidelines for Psychological Research. Code of Human Research Ethics. The British Psychological Society, 2011.
42. Yan, B., and Chen, G. AppJoy: personalized mobile application discovery. In *MobiSys '11*, ACM (June 2011), 113–126.