Privacy: What's different now?

Karen Spärck Jones
Computer Laboratory, University of Cambridge
William Gates Building, JJ Thomson Avenue, Cambridge CB3 0FD, England
sparckjones@cl.cam.ac.uk

This paper appears in Interdisciplinary Science Reviews, 28 (4), 2003 287-292.
It is based on talks at a British Academy Conversazione, May 2002, and at the British Association for the Advancement of Science, September 2002

Abstract: Modern computing technology makes it possible to record, preserve, collate, reconstruct and use detailed facts about individuals on a scale and in a style quite different from the past. With this technology, data is voluminous, not sparse; is permanent, not transient; and is detached, not tied to the individual's own observation. How do these developments change the reality, or the perception, of privacy for the individual? How, in particular, do they change the reality or perception of privacy for daily life, not just in legal situations or relations to authority? This paper examines the different properties of current computing technology, illustrating their consequences for personal data, and argues that these have profound implications for personal privacy and autonomy.

Introduction

This paper is about the impact of modern technology, and specifically information technology, on individual privacy, that is on the individual's actual or perceived privacy in relation to their personal autonomy. An individual's sense of themself depends crucially on being able to control their relations, in action or knowledge, with their environment and in particular their relations with other people.

Thus by *privacy* (for the individual) I mean not having things known about you that you don't choose to have known, or at least you know that they are known, and by whom.

One crucial point about privacy is thus that whether something is deemed private is not a property of the thing itself, but of whether you choose to regard it as private. That is, privacy is an extrinsic, not an intrinsic, property of a fact or piece of information about you. It is an independent point that many people choose to regard the same sort of information as private, or that what one might regard as especially intimate information like some internal body-scan image, or unique personal information like genetic data, has a special perceived status and hence might deserve particular legal protection. The *general* issue about privacy is rather different, as this paper is intended to show.

The other crucial point about privacy, in the fundamental view being taken here, is that knowing who knows about you carries with it some confidence, trust (or perhaps indifference)

as to what they do with what they know [5]. It does not necessarily matter what sort of people they are, or even that you may not know much about them as individuals.

Technology development has changed this. What sort of people know about you has become a pressing matter of principle; and the fact that many individuals, of whom you know absolutely nothing, may know about you and exploit their knowledge for any purpose, proper or improper, without your being aware of the fact, or giving your consent, has become a matter of concern. Privacy, or rather the lack of it, in its big government brother and commercial cousin aspects has become a serious, not merely fashionable, issue. Simon Davies, Director of Privacy International, said at a seminar in Cambridge in 2002 that the cause of personal privacy (in the UK) is already lost, through a combination of formal legislation and stealthy action, masked by apparent benevolence, that has taken advantage of its citizens' technical ignorance, as much as through their inertia or impotence.

But this paper is not about privacy in its big brother aspects, or in its institutional, formal and legal contexts. These depend, ultimately, on people's feeling about, and experience of, privacy as an essential constituent of their daily life as individuals. As this implies, privacy applies to mundane, ordinary and informal situations and relations, not only to commercial, institutional or formal ones. The object of this paper is to show how technology is affecting people's wholly ordinary and informal environments, and my concern is with the effect this has on the individual's actual privacy or, as importantly, their perception of privacy. What happens to privacy in these ordinary, informal and even friendly environments matters because it determines people's responses to the impact of technology in the more formal and unfriendly cases.

The paper is more about raising questions than answering them, mainly because technology is changing fast, and sociological data are lacking. My intention is to show what the properties of modern technology are, and what empirical (practical) consequences these have already had. My claim is that technology has already affected the actuality of privacy. The issue is whether it has changed the individual's perception of privacy, or the importance they assign to it.

The technology properties relating to data on individuals I will consider are those I have labelled

- permanence
- volume
- invisibility
- neutrality
- accessibility
- assembly
- remoteness

I will look first at how each of these separately bears on privacy and then, since these properties do not usually occur in isolation, at their impact when combined. This is where the rub really comes for the status and meaning of individual privacy.

I emphasise that "information technology" covers more than computing in the strict sense - it includes, for example, photography. But computing has been the critical agent of change,

both in its own right and in conjunction with other technologies like radio, as in mobile phones, and is central to information technology. So the technology properties with which I am concerned are those associated with computing, where the impact of computing on privacy has come from the extremely rapid development of computing power in terms of greater processing speed, increased storage capacity, wider communication connectivity and, also, lower machine and device size. This development is reflected in the appearance of terms like "ubiquitous" or "pervasive" computing (see [2,8]).

I distinguish privacy, referring to information about oneself, from *confidentiality*, which can refer to information about third parties; and also from *security*, which refers to the means by which privacy, or confidentiality, is assured, and which naturally includes computer security (for further detail on this see [1,3,6,7]. "Privacy" is used in the security literature to refer to the way in which some data item or transaction is kept private *within* a system. I am concerned here with the external view: human privacy in the face of systems. This is independent of whether system security is effective because, as I will show, the technology developments that are affecting privacy need not require any security, or at any rate nothing beyond preventing generic system failure.

Technology properties

1. Permanence, or persistence

By this I mean the permanent record of transient events and states. Permanence is familiar as writing and still photographs, and more recently as recorded sound and moving images. But computing has made it far easier to create and use records.

We can see this clearly with speech. It is easy to record speech as sound, but it is tedious in the extreme to find out from sound what people said, for example in order to find out whether Onora O'Neill talked about trust in banks one would have to listen to a half hour recording. Modern developments in speech recognition mean that we can transcribe speech as text and it is then orders of magnitude easier to search for words and content. So something essentially impermanent has now become permanent, with the new consequence that exactly what I said on some particular occasion last week can be quite easily recovered. [1]

The same applies to images, though the interpretive technology is much less powerful. It is possible to make and preserve very high-quality and detailed images, captured over very short time intervals. Digitising these images means it is easy to search large files of stills or video for specific low-level features or for images similar to a given one. It is not possible to search images by concept, e.g. pictures of women arranging flowers, but current technology can quite effectively select subfiles for human view.

The oldest forms of recording, writing and drawing, relate to time through the shorter or longer periods of time when they were conceived and produced. The important point about modern technology is that it makes it possible to capture much more transient, fine-grained data about events and states and their sequence in time, over long time spans; to record enormously much more such data than ever before; and to record it so that it can be searched.

[1] Transcription technology is far from perfect, but quite good enough in general to locate the bit of sound to listen to.

Permanence in this sense is independent of whether a particular technological medium is long life, or whether it is actually easy in practice to find something in a very large record base; and it is also independent of formal or even real deletability, i.e. whether only access pointers are removed or the actual medium is wiped.

This permanent data includes not only primary, but also secondary or associated, data. For example, for a perfectly ordinary activity like my preparing a document for public distribution, current software may keep a record of *all* the specific changes I have made throughout the whole process, regardless of whether I use the record or want it kept. The document editor is designed to be helpful in case I want to check or revert to an earlier state, but it is also keeping of record of my composition process, not just of what I may want kept as the final version.

2. Volume, or range

The rapid growth in the volume of data recorded for the individual is partly a consequence of the type of data recording just mentioned. But the spread of technology, especially information technology, has also stimulated data growth through the spread of point or transaction recording.

We are well used to the idea that significant sets of personal data are held by others e.g. national insurance office, vehicle licence bureau, bank, doctor, employer. We are also now well used to the fact that individual transactions, e.g. credit card purchases, are automatically recorded. The significant recent development is in the total quantity of data about the individual that is recorded and that is either personally identified or readily identifiable: for instance, me taking my bicycle from the rack outside my department (many times), dental visits, library borrowings, attendance at committee meetings, my order for a carpet, participating in a family anniversary. There is a huge mass of data about me that is entering the automated record(s) held by others. The existence of the mass is significant in itself, regardless of whether it is confidential, e.g. my doctor's file, or not, e.g. the department's web page, or in the hands of one, few, or many others.

This volume of data includes not only primary data, but also secondary or associated data of the kind illustrated in the previous section and, more generally, 'metadata' of all kinds, which may be very extensive. It may include, for instance, a log of all the accesses to a user's web page, or of pages visited by a user.

3. Invisibility

The metadata case draws attention to another current technology property: invisibility to the data subject.

Computer operations automatically generate internal administrative data, for instance that I tried to access a particular file (which I had good reason to believe I could legitimately read), or did access some other file (because its owner, my student, asked me to look at his project report). As a computer user I may in fact be able to access some machine metadata, for example see who has visited my web page. But in other cases, even if I could in principle check a machine log, say that recording my own activities, I would not necessarily be able

to interpret it. The data that operating systems gather for their own internal purposes are designed for their further program use, not for the people to whom they refer. But this system-created data for people may be very extensive and can be very informative, when unpacked, about what those people have been doing.

The same invisible but nontheless real data can also amplify machine records about me even when I am not a system user myself: for example the file with a scan of my broken ankle bone might be accessed by 50,000 people without my knowing this. This may be rather primitive or minimal as a fact about me, but so are many others, and it, like them, contributes to a very detailed picture of me that can be created without my knowing how detailed it is.

4. Neutrality, or 'indiscriminancy'

Modern technology can accept a huge variety of source data types, e.g. via scanning and digitisation: pretty well everything is grist to its bit mill.

As noted earlier for images, it does not follow that semantic content is easy to recover, though computer power means that we can work hard, and hence potentially effectively, at this using automated analysis and extraction tools. But the ease with which basic material can be acquired means that qualifying information that matters to me as a person - call it *my* pragmatic metadata - can be equally easily lost on the way. In itself technology cannot distinguish a public image from a private one, i.e. an image of such a public place as Trafalgar Square from an image of such a private place as the inside of my body taken with some latest medical device. Whether or not some image has some particular status, say as private, is normally not an integral part of an image but an attached tag, and everybody knows how easily objects and their labelling tags are parted.

Computer systems implement status mechanisms for files, defining who may access them, but only in a limit and formalised way. More substantive notions of privacy, refering to specific data content, are not easily implemented, and *within* files descriptive tags or comments that seek to capture this can be separated from the data to which they apply. Parts of files, for instance, paragraphs from a text, extracts from a message, can be selected and passed elsewhere with their original context and status lost, but may still be readily recognised and associated with the individual to whom they belong. The longer the chain of copy or extract, the more the 'neutral' status.

Even though people vary in what they regard as intimate and might thus want to keep private, for example with respect to their bodies, emotions or personal relations, it does not follow that they will be happy to have information about such things as personally identified or identifiable data points become mere items in an autonomous automated system.

5. Accessibility, or availability

Data about an individual being accessible is not just a natural consequence of, say, my having a web page anyone on the internet can reach, which follows a conscious decision on my part. Accessibility is much more a byproduct of the many valuable or harmless things people do anyway, multiplied by the ease of copying and by endemic natural leakage. Thus, illustrating the inavoidable, less desirable consequences that things we otherwise want can

have, organisations as a matter of course install anti-virus software that can only work by inspecting every bit in incoming data streams including, therefore, messages to me. Virtually anything in any record files anywhere, whether on the Web or in other places, can be read by some number of people, small or large, authorised or unauthorised, including much that is naturally, or necessarily, not encrypted. Electronic versions of newspapers are naturally posted in clear, and this makes a lot of information about many individuals very easy to reach, and in a form that make it very easy to pick up and re-use.

Perhaps more importantly, the ease with which we can copy electronically increases data accessibility by reducing search effort. (Computers and communications support single data locations, but people still logically package data in multiple ways, so the data is accessible in multiple ways regardless of whether the end object is at a single location.) Even at the local level, copiers and computers mean that data on or associated with an individual spreads to more other people, even if not generally. But the most striking spread is via the Web, for the best motives: genealogical and family material is enthusiastically promulgated by those interested in learning about, and making contact with, family members.

6. Assembly, or constitution

Computing technology makes it enormously much easier to assemble data from different sources.

There are many effective tools for searching for and collecting items from many quite separate and scattered points. Even if there is no novelty in the component items, combination gives a richer view of the individual through the way it establishes relations between items and through the wide variety of items that can be brought together. For instance, a quite superficial search on the name of a member of my department, who does not have a web page or publications, pulled out a surprising set of items illustrating roles and relations not obvious from her current formal position - surprising both as an assembly and in others' view of the person concerned. Some items were not accurate, but that also holds for non-technology data. The important point is that because so much data is now recorded, and however scattered it is, we can pull together an overall very detailed and multi-facetted package on an individual.

Moreover the ease with which material can be edited - text reworded or extracted, images cropped, distinct originals cut and pasted together, etc, leads readily to new data items that are in turn recorded and propagated as records, with all the supposed objectivity that automated records (like printed packages) are likely to acquire. None of this may be improperly intended, the changes may be fully justified for all kinds of good reasons, but the outcome is nevertheless new data items about an individual.

7. Remoteness, or anonymity

This final technology property is implicit in the previous ones, but deserves explicit discussion in its own right. By remoteness I mean the fact that data about me are open to others who are not only physically far away from me, but are logically so. Data about me can be accessed and used by people of whom I know nothing either as individuals or classes.

In the past we in general knew, or knew of, the people who knew us, or at least knew

the classes of people who would know significant amounts about us, from family through neighbours to unknown individuals but known types e.g. in a hospital or bank. We knew enough about the links in the chain to 'personalise' the connection, especially for substantive information, or we could be reasonably confident that data about us obtained by a passerby in the street, or learnt in gossip in a remote bar, would be just fragments and unlikely to return to affect us. This was true despite, for example, ubiquitous police spies in Schubert's Vienna: they were visible people (say cafe waiters) even if not identified as spies, or of known type (like the local bureaucrats).

Now with technology we may have no connection at all with people who can access large amounts of data about us and can use it as they wish, as if the individual they are constructing is you; and in general we have no means of discovering who these people are and hence of making a connection that has some personality to it.

I emphasise again that there need not be any big brother elements - authoritarian or improper - in this at all. For example, my medical data could be exploited for epidemiological or sociological studies I know nothing whatever about, and about which I might in particular cases be concerned even if I have given my general consent to my data being used for such purposes. Or your nine-year old daughter's prize essay, posted on her school web pages and intended for its community readers, could be taken in a rather different way by some faceless bureaucrat as evidence of the school's low literacy standards.

Combination

Each of the seven features in itself affects privacy. But their effect in combination is much greater, as even a tiny example can show.

Thus I discovered, by a simple web search on my grandmother's maiden name (which is also one of my Christian names), a web page from a newsletter with the text of remarks about me that I had heard delivered on a particular occasion but did not know had even been printed, let alone published electronically, illustrated by a photograph that the author of the remarks had not provided and I had never seen before. The newsletter itself also placed me in a new institutional context, as associated, if only informally, with an organisation that had no connection with the original occasion and with which I myself had no current connection.

There is nothing particularly pernicious about this, or anything very different as one small case from analogues in the past. Its importance is in the fact that things like this happen so easily, and so much and also, crucially, not knowingly as far as one's own decision or at least awareness is concerned.

Assessment

So, going beyond the anecdotal, what do the technology factors add up to ? Do they support my claim that they have changed personal privacy in our society and, specifically, that privacy has been undermined?

First, many data *items* items are very weak, even those I provide directly myself: my title for a talk to be advertised, for example, or of my preferred holiday country for travel offers, do not in themselves say much about me. Many items may be even less informative, and not very reliably personalised. This is true whether or not I was directly reponsible for the item

or it was only a byproduct of something else I did. For instance, the fact may only be that there was an AltaVista search on "playing pontoon" from my machine, or that my phone number was logged as calling by some other phone. A single search on "playing pontoon", or calling a phone number once, does not say much about anyone; and these items may not in reality apply to me personally, since someone else could use my computer or my phone.

Second, even where I am not directly responsible for the item, as I would be if I had posted a brief biography on my web page, data items may follow from information about me which is public, e.g. my phone number; or I may be aware that they are produced, for example because I know that AltaVista searches are logged, or that I am observed by CCTV in the street. These 'monitor' items may not seem to be different from traditional forms of observation: what is different about a till record of my buying rice crispies on my credit card when the person next to me in the checkout queue can already see them in my basket?

Third, all of these data items can be a direct consequence of my using computers for purposes that I think good, and may even be used to support those purposes, in a positive feedback loop. People like being able to post digitised family photographs on their web sites and send them round by email, or being able to organise holidays online. The same holds where people are not computer users themselves, but are quite happy to have others use data about them, for example to advertise their house for sale, or to check their medication. Again, it is my choice to do an AltaVista search, and I can hardly complain if search engines seek to improve their performance, which may benefit me, by analysing user search logs.

In other words, people can and do welcome the technology, and are clearly willing to exchange privacy for other benefits, for example to trade detailed purchase records for the convenience of having credit cards. Yet further, why does the combination potential matter when so many people are willing to complete amazingly extensive and intimate marketing questionnaires?

Thus contrary to my claim that privacy has been damagingly reduced, it can be maintained that data is not information, especially when many items are minimal or dispersed or are public anyway. Or alternatively, maintained that even if a sufficient number of individually weak data items do add up to quite a lot of quite personal information about an individual, people are willing to accept some loss of privacy in return for what they gain from the technology.

My argument is that the ramifications and implications of technology are far greater than is realised, and specifically that the quantitative growth in data recording is certainly leading to a qualitative change in the actuality of privacy. Massive data is in itself information, but far more importantly, it is the use that can be, and is being made, of information that matters. The quantitative growth is at an ever finer grain, and increasingly the product of automatic generators some distance from the individual's original action. For example, radio frequency identification (RFID) technology allows unique object tracking, and there are already references to 'the internet of things'. But if tracking chips are embedded in objects so we cannot remove them without destroying the objects, and we do not have any practical choice about buying the objects themselves, we are stuck with the record of what we bought, when we bought it, and where we have taken it since.

Thus while I have concentrated in this paper on the immediate and mundane manifestations of information technology, and not on its threatening potential, most of the rapidly growing concern about the technology's impact on personal privacy is about the proper and improper uses that authorities may make of knowledge about individuals, and in particular of knowledge about us that has hitherto been regarded as sufficiently private. The ease with

which, by commission or omission, technology can be used to subvert, or simply bypass, privacy is also a criminal threat. Technology countermeasures, however powerful - as for example quantum cryptography might be, cannot prevent this. They can reduce risk but not eliminate it entirely, because systems exist for human purposes and so at some point are vulnerable to human weakness. Human beings are necessarily involved in computer systems, and they are no more careful, reliable or honest than they have ever been. Thus I may transmit my message quite securely under the latest encryption technology, but what it says can end up in an unprotected database [4].

Seeing the impact of technology on privacy primarily as an 'us-them' problem is not, however important this is, my main concern. My argument is that the impact technology has already had on privacy is at a much more fundamental, because ordinary and everyday, level.

Clearly the *perception* of privacy cannot change if people are ignorant of what is happening, and it won't change if people deliberately ignore the situation. But the belief that one's view of personal privacy needn't change, though there is a mass of material out there, because other people will not do anything with it, or because what they do with it won't or needn't affect one or matter to one, or because whatever they do it doesn't matter if I don't know about it, is not a rational belief. Other people will do things with the data they have about me, and this will affect me, if only through the way it leads others to see me. In particular, there is no reason to suppose that data will remain scattered, and therefore that the development of information technology does not imply any fundamentally new problem about privacy.

It is also not self-evident that technology need not have any effect either on my actual privacy or on my perception of it, because *real* privacy is in the mind and no-one knows what I think. However much external data, even correct facts, are known about me, we may suppose that they cannot affect the sovereignty of my own thinking about myself. But as those with the experience of totalitarian regimes can testify, this is a tough line to hold.

Thus I do want to ask whether privacy and hence my feeling of myself as an individual is being undermined, not so much by a growing lack of control over what is known about me as by the fact that I don't know who knows it. This loss of control is quite involuntary: trying to combat it would require a degree of isolation analogous, in everyday life, to not speaking to anyone or not even walking down the street. One might think that people in the past lacked privacy, whether through overcrowding in houses, or rootedness in villages, or the presence of servants, or the requirements for religious conformity, so things are not really different now. I want to suggest that technology has, or is rapidly, changing the reality of privacy, primarily through our lack of knowledge about what others see and know about us, and about how much they see and know, how they can reconstruct us from our bits. This is quite different from the past.

It may, perhaps, ultimately be a philosophical question whether, without any privacy at all, one can be a person, because one has no autonomy through control of how one appears or relates to others. But we all have a (our own) concept of privacy, which we may invoke even against out nearest and dearest, regardless of what the law says, or whether it deals only with media intrusion on the great. Current technology developments are so rapid and so far reaching that even their unphilosophical effects on our privacy and our perceptions of this will be profound; and even though these changes are inevitable, and in many ways beneficial, it does not follow that we should sleepwalk through their implications. In particular, it is only by thinking about the way they work for us in quite harmless contexts, that we can seek to guard against the way they can be made to work against us in harmful ones.

References

1. Anderson, R. *Security engineering*, New York: Wiley, 2001.
2. Beckwith, R. 'Designing for ubiquity: the perception of privacy', *Pervasive Computing*, April-June 2003, 40-46.
3. IEEE Computer Society Symposia on Security and Privacy.
4. Needham, R.M. 'Computer security?', *Philosophical Transactions of the Royal Society, Series A*, 361, 2003, 1549-1555.
5. O.Neill, O. 'A question of trust. The Reith Lectures 2002', Cambridge, Cambridge University Press, 2002.
6. Serjeantov, A., Dingledine, R. and Syverson, P.F. (Eds.) *Information hiding*, LNCS 2578, Berlin: Springer, 2003.
7. Spärck Jones, K. 'Computer security - a layperson's guide, from the bottom up', Technical Report 550, Computer Laboratory, University of Cambridge, 2002.
8. Stajano, F. and Crowcroft, J. 'The butt of the iceberg: hidden security problems of ubiquitous systems', in *Ambient intelligence: impact on embedded-system design*, Ed. A.A. Basten, M. Geilen and H. de Groot, Dordrecht: Kluwer, 2003.