Who am i & why should you (or i) care?

Who am I? I am a person defined by a collection of relationships to other people, and their understanding of my various behavours with respect to them.

Jackie Chan the philosopher, martial artist, buffoon, film actor and comedian, in his film of this name, loses his memory at the start of the plot He still has skills (he can walk, talk, carry out wonderful and humorous fights and stunts) but he can't remember his name – falling in to the hands of a primitive rural society who do not speak much English (or Cantonese), he asks them "Who am I", and becomes known by them as "*whoami*". The salient point, however, is that despite his amnesia, his behaviour, and his essential character, despite lacking an apparent identity, behaves just like the typical Jackie Chan hero, humorously and quintessentially morally. Thus the set of new relationships with new people resemble his old. The *handle* for his character is irrelevant. We'll come back to this when talking about identity and identifiers, but writing, as I do, in England, I'd like to point out the nearly unique right we have to not be identified (or re-identified) via a single unique handle – we can choose to separate out groups of relationships with subsets of friends and agencies by using different persona (as is natural) and different handles. I myself do this routinely as my name at birth was Jonathan, but only my immediate family (my kinship group) call me that. Friends from when I was a child up until I first went to University (i.e. close friends of thirty years or more) call me Johnny. Since then, people I have met as colleagues, or made friends with or mere acquaintances call me Jon (not John☺. This is a convenient shorthand reminder when someone phones or emails me, just which class of relationship I have with the originator. (It also catches a large class of spammers and *phishers* who have gotten my handle from some official document or other from the Interweb, and so use "Jonathan" which immediately triggers precisely the right cognitive dissonance between their offer of performance enhancing drugs or wonderful African inheritance tales, and the type of person I am thinking of making these fine goods and services available to me.

The rise of the robots

In discussing identity and relationships in the 21st century, it behoves us to discuss synthetic beings too. I am fairly confident that as well as AIs (and Siri), autonomous robots will walk (or at least drive) the streets in our every day lives quite soon now. Philosophers and ethical thinkers have been considering the consequences for a surprisingly long time. The legend and tale of the Golem & the Rabbi of Prague, with its primitive on/off switch (the Hebrew words for death/life) turned into a lethal killer, perhaps due to its lack of soul. This inspired not just Carol Kapec's play RUR (the first use of the work-a-day word, robot an artificial life, but probably Mary Shelley's acute dissection of the existential angst of soul-less Baron Frankenstein's monster – note though, despite having no soul, the creature os capable of sophisticated moral thought and engages in a long debate (most people skip this part of the book) with its creator. Much later, the scientist and hard-SF writer, Isaac Asimov had much fun with stories around what could go wrong with robots "hardwired" with three laws (and later, four) Many SF satirists had fun with variances of Asimov's laws – most amusingly was John Sladek's tale of the robot which, when turned on, immediately starts gaily slaughtering humans all about him – when caught and asked "why", he replies "oh, I misread the law – I thought it said "A robot shall not inure a human"

This anecdote isn't here pointlessly but illustrates that just as humans have failings, artificial beings will be the subject of programming errors too. Just because something is controlled by synthetic rules doesn't mean it will behave well. Indeed, this is trivially obvious with gadgets that can be weapons or plough-shared peacefully. This means that there are additional hidden relationships with a synthetic being, concerning the provenance, and possible liability to be born by someone else when misbehaviour occurs. In the weapons world, explosive manufactures (e.g. making Semtex under license (those bouncing Czechs again) , watermark the materials so when misused, an audit trail of who sold what to the bad guys inappropriately is available. We could do this with software (and drones) too.

Of course there are limits to what might be achieved, but recent work on tracking down *botnets* for hire has centered on following the money (UC Berkeley's work on click trajectory). This works fairly well for normal criminals (ones who are rationally motivated and for some reason want to make money doing stuff that isn't allowed). However, it will fail spectacularly for mad people, however, who follow their god rather than Mammon. Beware the hegemonizing swarm, as the saying goes.

Who cares?

Why do people care so much who I really am? Why should anyone outside of actual family, circle of friends, colleagues and acquaintances want access to anything about my relationships and behaviour. I'd say that there are at least three categories of invaders of a personal space.

1. Prurient public interest in celebrities, even minor celebs – even members of the great unwashed public who happened to be having their McLuhan moment A large part of the press (and TV) is taken up with propagating gossip about people for whom the vast majority of readers or audience have absolutely no personal links, nor ever will have. Thus people have been provided with salacious gossip, by journalists (especially in the UK, recently) tempted to go beyond all reasonable limits in hunting down incredible levels of spurious detail about individuals caught in the limelight.

   Who really was present at Marilyn Monroe's death? I really don't care but a lot of trees died to print a lot of stuff about it so someone does. Who *really* was she. (Again for English readers, go look up stories about what Amy Winehouse was really like).

   You know what this sort of behaviour reminds me of? Two things – the pointless and evil laying bare a person's innermost self that is what is done by torturers and interrogators (especially brainwashing). It is, essentially, a form of pornography.

2. The second class of space invaders are the two-faced market driven Interweb businesses. Why should Google or Facebook have any commercial interest in little 'ole me? Basically for two reasons:

   They want to make money delivering targeted adverts and recommendations to me – the better targeted, the more likely I am to act, the more likely their advertiser (or the business selling the goods or serveices being advertised) will

cough up. Measuring click through is not a precise business, and indeed, some grey businesses generate *click-fraud* (often hiring those bots again:) – this is no different than self citing (a bad old academic practice to boost an author's papers' citation counts)

Secondly, the aggregators want to run analytics (basically, market research) to determine the statistics of interest in goods and services amongst various classers of user.

The organizations doing this are not evil, per se – they are driven by profit motives. They observe that the fine and coarse grain data above is vastly superior, faster and cheaper to gather than traditional market research information (just as social scientists have discived that he interweb and smart phones make it much easier to do really big studies of humans).

Of course, it is only a matter of time before a really huge leak of personal data from one of these outfits happens (perhaps due to disgrunteled, mad, threatened or bribed employee, or perhaps just due to plain old bugs (see mad robots again).

3. The third class of user is the newest, and that is the Government who are here en masse once more to help us, by exploiting all that *Big Data* that their various minions have been accumulating about all of us.

There are ostensibly good reasons for a government to do this – for example evidence based health, and tuning energy and other policies might all be better done if the data about our unhealthy and un-green life styles is shared with as many outfits as you can shake a stick at. I'm not convinced the benefits outweigh the risks. Aggregate health statistics are usually good enough – our health protection agencies are being cut, so that doesn't indicate good intentions – monetizing the data set value by letting private interests get at it is strongly counter indicated with the invasion of privacy we see with online services (whether the data per person is relatively low value compared to health or energy or driving practice records).

A second use of Big Data by government agencies is the job the various cybersecurity groups get up to in finding the bad guys (odd looking cliques in social graph). I am in principle ok with strictly controlled use of such metadata for intelligence agencies doing this for, err, intelligence, but I'd be really upset of it was done routinely for evidence for policing – the effect would be to put us all in a *panopticon*. There would inevitably be mission creep (councils looking at who sees who, to guess whether people are eligible for this school or that, this rubbish collection service or that) – This would be toxic.

Is the past 12 years "typical", representative of the future we can expect, or is it just a brief mad interlude ? I'm not at all sure that it is impossible to think of alternative worlds where we have a large number of small decentralized systems for managing the only version of our relationships and behaviour with only the set of people we care about – the infrastructure for such systems would be extremely cheap (so cheap

as to make the data plan on your smart phone dwarf it) – indeed, we already have systems like this (my kids schools have web sites and mail lists which are closed to just the parents, teachers and children of the school. My local sports club has the same.

Do we (e.g. wearing an investor's hat) really want to base a future on a small window of experience with the two-faced market of the 2000s? Just think of some poor person with some spare cash in England in 1780 being advised to "buy canals, young boy". 20 years later they were obsoeted by trains. Some people think you can't tip the market or move users -- the first mover's advantage – but these guys weren't first- Altavista (and many other tools) predate them. Note the company that built altavista (DEC) no longer even exists. Note Usenet news (a system which closely ressembles a halfway house between Twitter and Facebook) lasted 30 years and is no more. Sadly, few people remember it (Dear Reader, would you hazard a guess how often Godwin's law is re-discovered in new media each year?).

All your life is online, and on a handful of near monopoly services. If they make a mistake and release it, it's a one shot disaster. If they go broke, and someone else decides to acquire their servers and disks, and to change the business model on you, it's a catastrophe. Something has to change.

Social Work

The Oxford Anthropologist, Robin Dunbar is not just a number (Dunbar's number, the size of a hunter gather group, or a unit of the Gore Tex company or a convenient army size gang, is 150). This work was part of a much larger endeavour to look at the structure and causes of social networks, originally just in the real world. Dunbar's group found that primates all share similar social groups, bonded together by relationships and gossip (yes, Chimpanzees gossip – they do it while grooming each other) – the larger the brain, the larger the social group. Human's bound large groups with large brains, using the power of language and abstraction. We tell each other stories about one another. We can cope with stories with very large numbers of levels of indirection (Shakespeare tells us that Iago lies to Othello about Desdemona having an affair – we can cope with 6 levels of indirection, separation or even anxiety.

The closer (and most recent and frequent) interactions with a given person re-enforce our model of that person. Our theory of mind (our ability to put ourselves in someone else's shoes ("if I was him, I'd phone her up", or "I can't believe she just said that to her seriously", etc, etc), combined with language is a powerful system for building, binding and maintaining society.

People at the periphery of our social group (acquaintances, people we have weak ties with – our friend's friends, who we see only occasionally, or in superficial contexts have the least links and the weakest position in our cognitive model of the group. We *grok* them least. We are most surprised by them, and most astonished by their actions. We probably trust them least. (Note this doesn't mean we like the closest people – we might hate them, but we have the most accurate model of what they might do in a given hypothetical situation). We can also explain our model to people (using that wonderful tool for abstraction, language). The strongest bonds are re-enforced by

quite simple shared experience – dancing, drinking, laughing and so on. Organisations that want people to bond know this too, hence the eating and drinking clubs of the ruling classes, and the training methods of armies.

We (and other people) have studied whether these laws (size of group, levels of relationship) are as true in online social networks as in Real Life (or RL as gamers and hackers refer to it). And we find that the numbers are in fact very similar – you might have 1500 "friends" on Facebook, but the ones you actually have similar likes and, crucially, interact with regularly are probably in 6 layers and number around 150, on average.

These groups are not static, hence, for example when we move somewhere else in the world (for college or work) we make new friends and, importantly, we start to forget old ones. We have to put in quite a bit of energy to keep up long distance relationships, although the occasional meeting can refresh a deep friendship for a remarkably long time.

In the online world, then, with the technology they have developed, the companies that have potential full capture of our lives are in a uniquely weird place. Instead of being like a bank, or our GP or a teacher who has a specific narrow and controlled view of data from a single domain (money, health, education) concerning us an is therefore completely peripheral to our social net, these two-sided marketers are actually in level 0 of our net – the very center. This is most bizarre. In some senses, this reminds me of people on the autism spectrum who wander into the midst of a social circle and disrupt things due to their lack of empathy, or any theory of mind (like the synthetic beings ("replicants") in Phillip K. Dick's "Do Androids Dream of Electric Sheep") - are cloud companies just high functioning Aspergers with added stalker syndromes?

Can The Technology (and business model) change?

Yes, I think the technology can change.

Firstly, as I mentioned, the cost of the storage and computing resource to support our own social net is remarkably small – much less than the most of the devices and services we use to access it. Certainly today's Cloud services run on very big hardware – but a lot of that is because they have to run the expensive analytics (market research) software (map/reduce or Hadoop or bulk synchronous jobs run on 10s of thousands of Facebook computers in their data centers upwards of 3000 times a day). None of that is necessary of we just want our data on a secure personal cloud that we pay (a few euros a month) for.

However, we can even continue to support the two-faced *marketroids* if we really must, without massive alien invasion of our personal space (accidental or malicious). Two techniques are making some progress in terms of being affordable ways of keeping information private but extracting some statistical results from collections of data. One is the technique called *differential privacy* which computes a rate of information leakage from a query over a database (e.g. collection of a number of users profiles or likes). The other (still computationally expensive) is a technique called

homomorphic cryptography which allows computation over encrypted data (without decryption first).

These an be used to build systems that are privacy-preserving, for functions like search, advertising, and even analytics.

A interesting consequence of designing a system with good privacy properties is that it offers possibility of users "giving" *more* data, actually permitting better targeting of ads (without the advertising potentially knowing to whom the ad went) and better analytics – What is not to like for any of the stakeholders in that scenario?
Two alternatives also suggest themselves.

The first might be a decentralized system where resources are shared between users. There are no providers except the users themselves. Such a system might operate lie filesharing systems of the past (most notably, Bittorrent) with a tit-for-tat resource sharing enforcement mechanism, so that if you want to be in the club, you have to share too. The arguments against decentralized systems are mainly (mainly concerned with how to provide high availability and low latency in response to requests for information). As more and more users (at least in the developed world) have fiber optic net access and 4G is deploying the uplink performance impact on latency is rapidly decreasing. However, reliability is still an issue. Some folks have proposed a hybrid, with a pay-per-use central store just for backup (Skype use supernodes for this purpose too). This removes the potential temptation for the central service provider (a.k.a. "honest, but curious") to look at my data, since it only holds an encrypted backup. I can replicate this as much as I want, and it doesn't have to be on superfast servers, just there for the occasional system failure. I can replicate frequent and recently accessed content of my own on as many of my friends peer-machines as I choose to trust (hey that is the data they are looking at about me anyhow, so what's not to trust?). Hence one would have a decentralized online social network, but a federated centralized backup store for it.

A completely different idea might be to build a system that has no privacy technology or anonymity at all, but enforces true names & identities – we might christen such a system *Paradoxy*. In *Paradoxy*, all gossip is visible, so why would people behave maliciously? In RL, most people don't, because of the social sanctions, so perhaps in *Paradoxy*, social capital and pressure would be sufficient to encourage good, and discourage bad behaviour? Such a system would be interesting to try out to see what norms emerged.

A Futurist Policy Manifesto

If technology can be build that makes this feasible, should not Privacy Law change to require it? - Make It So!

My mantra is: Only hold data that is pertinent, for so long as relevant

My long term goal is to go further - don't hold data at all.

- I "hold' my data.
- I give you a capability to ask me for my data for so long as I allow...
- Audit trail tells me who looked at it when.

- Now no need for one single identity (which is an illusion anyhow)

- Me jon(a(than) = work, friends, close family

- My kids - two last names = parents, nationilty

- Future - same as past (but not present) -

Exploit unique UK position

- 1 id per relationship: bank, Tescos, Amazon, doctor, school work, friends, family with associated keys to data
- Data owned by me (replicated encrypted in a million clouds)
- No aggregation allowed by others (only me:)

Consequence of technology + and legal shift

- Allows +me+ to monetize my person

- Tell how much value my store loyalty card is worth

- Provenance - digital footprint/breadcrumbs

- can track s/w

- and robots (or more importantly their programmers or priests)

- and AIs too

Who pays?

I do - because it costs peanuts, and I am not a chimpanzee.

The data (to be added)

- Total Facebook or Google revenue/number of users (3-5$ per year)

- Total storage per user on Facebook or Google per user? (<100Gbytes per person)

Subscription instead of Panopticon

As with Music, film and other high quality content already heading that way (compare iTunes, Last.fm, Spotify)

Note celebreties (after all, who am i) aren't on Facebook...possibly high paying customers

Do you really want to be low quality, marginal profit, product?

Acknowledgements